НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ ЗАХИСТУ ІНФОРМАЦІЇ

«До захисту допущено»
В.о. завідувача кафедрою
_____ М.М.Савчук
(підпис) (ініціали, прізвище)
“___ ”_____2019 р.

Дипломна робота

на здобуття ступеня бакалавра

з напряму підготовки: 6.040301 «Прикладна математика»
(код і назва)

на тему: Класифікація перестановок зі спеціальними властивостями та оцінка потужності класів

Виконав (-ла): студент (-ка) 4 курсу, групи ФІ-52
(шифр групи)

Бурлака Марія Костянтинівна _____
(прізвище, ім’я, по батькові) (підпис)

Керівник: доктор фіз.-мат. наук, в. о. завідувача кафедри ММЗІ ФТІ
НТУУ “КПІ ім. Ігоря Сікорського” Савчук М. М. _____ _____
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Консультант: _____ _____ _____
(назва розділу) (посада, вчене звання, науковий ступінь, прізвище, ініціали) (підпис)

Рецензент: професор кафедри інформаційної безпеки, доктор точних
наук Качинський Анатолій Броніславович _____
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали) (підпис)

Засвідчую, що у цій дипломній роботі немає запозичень
з праць інших авторів без відповідних посилань.
Студент _____
(підпис)

Національний технічний університет України
«Київський політехнічний інститут
імені Ігоря Сікорського»
Фізико-технічний інститут

Кафедра математичних методів захисту інформації

Рівень вищої освіти – перший (бакалаврський)

Напрям підготовки - 6.040301 «Прикладна математика»

ЗАТВЕРДЖУЮ
В.о. завідувача кафедрою

М.М.Савчук

_____
(підпис) (ініціали, прізвище)

«___»_____20__ р.

ЗАВДАННЯ
на дипломну роботу студенту

Бурлака Марія Костянтинівна
(прізвище, ім'я, по батькові)

1. Тема роботи Класифікація перестановок зі спеціальними властивостями та оцінка потужності класів,
керівник роботи доктор фіз.-мат. наук, в. о. завідувача кафедри ММЗІ ФТІ НТУУ "КПІ ім. Ігоря Сікорського" Савчук М. М.,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
затверджені наказом по університету від _____ р. № _____

2. Термін подання студентом роботи _____

3. Вихідні дані до роботи _____
_____
_____
_____

4. Зміст роботи визначення застосовності перестановки у якості підключа шифрування у роторних машинах за допомогою аналізу криптографічних властивостей перестановки і оцінка кількості перестановок із заданими характеристиками,

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо) _____

_____
_____
_____

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання _____

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1.	провести огляд опублікованих джерел за тематикою дослідження	31.08.18	
2.	дослідити вплив характеристики перестановки на результат зашифрування	30.11.18	
3.	створити програму для реалізації розбиття множини перестановок на класи	31.03.19	
4.	оцінити потужності отриманих класів	20.04.19	
5.	перевірити якість оцінки потужності	10.05.19	

Студент _____ Бурлака М. К.
(підпис) (ініціали, прізвище)

Керівник роботи _____ Савчук М. Н.

РЕФЕРАТ

Кваліфікаційна робота містить: 71 стор., 1 рисунок, 11 таблиць, 13 джерел.

Метою дослідження є класифікація підстановок в ключах роторних шифрувальних машин в залежності від їх криптографічних характеристик, експериментальне отримання статистичних оцінок потужностей класів для підстановок різного розміру, порівняння точності оцінок при використанні різних апроксимацій для ймовірнісних розподілів в методі Монте-Карло.

Об'єктом дослідження є ключі зашифрування у роторних машинах.

Предметом дослідження є криптографічні характеристики підстановок над алфавітами.

В ході роботи розроблено алгоритми для побудови класів підстановок за характеристиками методом повного перебору перестановок та методом статистичного моделювання Монте-Карло. Виділено класи підстановок довжин 11, 26, 30, 31, 32, 33, 45 та 55, побудовано довірчі інтервали для потужностей отриманих класів з використанням різних методів. Проведено аналіз результатів, отриманих в результаті апроксимацій для ймовірнісних розподілів в методі Монте-Карло.

СИМЕТРИЧНА КРИПТОГРАФІЯ, ПЕРЕСТАНОВКА, ХАРАКТЕРИСТИКА, ДОВІРЧЕ ОЦІНЮВАННЯ, ДОВІРЧИЙ ІНТЕРВАЛ, ТОЧКОВА ОЦІНКА, РОТОРНИЙ ШИФРАТОР

ABSTRACT

The thesis contains: 71 p., 1 figure, 11 tables, 13 references.

The aim of the research is to classify permutations in the keys of rotary cryptographic machines depending on their cryptographic characteristics, to experimentally obtain statistical estimates of the power classes for permutations of different lengths, to compare the accuracy of the estimates using different approximations for the probability distributions in the Monte Carlo method.

The object is cipher keys in rotor engines.

The subject is cryptographic properties of alphabets permutations.

During the research, algorithms were developed for constructing classes of permutations according to characteristics by the method of complete enumeration of permutations and the method of statistical simulation of Monte Carlo. The classes of permutations of lengths 11, 26, 30, 31, 32, 33, 45, 55 are allocated, the confidence intervals for the powers of the obtained classes are constructed using various methods. The analysis of the results obtained as a result of approximations for the probability distributions in the Monte Carlo method has been carried out.

SYMMETRIC CRYPTOGRAPHY, PERMUTATION, CHARACTERISTICS, CONFIDENCE ESTIMATION, CONFIDENCE INTERVAL, POINT ESTIMATION, ROTARY ENCRYPTION

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

$\bar{A}$ — доповнення множини A

$P(A)$ — імовірність події A

$\boxplus$ — операція суми за модулем

ВСТУП

Актуальність дослідження. Для передачі секретної інформації під час Другої світової війни широко використовувалися роторні шифратори, зокрема, різні модифікації Енігми, які використовують перестановки на алфавіті у якості ключа зашифрування. Проте і сьогодні роторні шифратори не втратили своєї актуальності. Досі вони використовуються для збереження чутливої інформації у багатьох країнах як у воєнних структурах, так і ентузіастами.

Метою дослідження є класифікація підстановок в ключах роторних шифрувальних машин в залежності від їх криптографічних характеристик, експериментальне отримання статистичних оцінок потужностей класів для підстановок різного розміру, порівняння точності оцінок при використанні різних апроксимацій для ймовірнісних розподілів в методі Монте-Карло. Для досягнення мети необхідно розв'язати **задачу дослідження**, яка полягає у визначенні застосовності перестановки у якості підключа шифрування у роторних машинах за допомогою аналізу криптографічних властивостей перестановки і оцінці кількості перестановок із заданими характеристиками. Для розв'язання задачі необхідно вирішити такі завдання:

1) провести огляд опублікованих джерел за тематикою дослідження;

2) дослідити вплив характеристики перестановки на результат зашифрування;

3) створити програму для реалізації розбиття множини перестановок на класи;

4) оцінити потужності отриманих класів;

5) перевірити якість оцінки потужності.

Об'єктом дослідження є ключі зашифрування у роторних машинах.

Предметом дослідження є криптографічні характеристики підстановок над алфавітами.

При розв'язанні поставлених завдань використовувались наступні *методи дослідження*: теорія імовірностей, математична статистика, метод статистичного моделювання Монте-Карло, криптоаналіз, алгебраїчні та машинні експерименти.

Наукова новизна отриманих результатів полягає у вперше отриманому розбитті перестановок різних довжин на класи за виглядом їх характеристик; оцінці потужностей запропонованих класів.

Практичне значення результатів полягає у використанні побудованих таблиць з характеристиками перестановок при виборі ключа зашифрування роторної машини. Дани таблиці дають змогу підвищити якість результату роботи шифратора.

Апробація результатів та публікації. Результати роботи були опубліковані на Всеукраїнській науково-практичній конференції студентів, аспірантів та молодих вчених на базі Фізико-технічного інституту Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського" у напрямку «Теоретичні та прикладні проблеми криптографічного захисту інформації».

1 ТЕОРЕТИЧНІ ВІДОМОСТІ

Даний розділ містить означення і поняття, необхідні для подальшого розуміння і використання в дослідженні, наведені основні матеріали, пов'язані з шифрувальною машиною Енігма та принципом її роботи, розглянуті криптоаналіз машини й використовувані у ній перестановки. Також розглянуто метод статистичних випробувань (метод Монте-Карло) і його застосування в дослідженні.

1.1 Енігма та її історія

Захист інформації від сторонніх людей завжди був актуальним питанням. Протягом усієї історії людства створювалися і вдосконалювалися різноманітні схеми, системи і засоби, що дозволяли передавати чутливі данні з меншим ризиком їх розкриття. Одним із таких засобів є роторні шифрувальні машини, зокрема, Енігма - переносна роторна шифрувальна машина. Такі шифрувальні машини використовували для зашифрування та розшифрування секретних повідомлень починаючи з 20-х років XX сторіччя.

Енігма використовувалася як у воєнних, так і у комерційних цілях, проте найбільшого застосування Енігма набула у нацистській Германії під час Другої Світової війни. Перші спроби зламати код Енігми були здійснені польськими дешифрувальниками. У міжвоєнні роки там сформувалася команда математиків, що працювала над задачею спрощення і пришвидшення процесу розшифровування шифрів Енігми, проте досягти цього змогли тільки після 1940 року, коли роботу очолили інженери Бретчлі-парку в Англії.

З точки зору сучасної криптографії шифр Енігми є досить слабким, але лише збіг таких обставин, як надзвичайно сильна команда криптоаналітиків, помилки операторів, завчасно відомий текст повідомлень (у тому числі метеопрогнози) та перехват робочих моделей машини, дав можливість розшифровувати і читати повідомлення.

Значення роторних шифрувальних машин, зокрема, Енігми у сучасних умовах полягає в тому, що вони дали поштовх на розвиток загальновживаного сьогодні шифру DES (Data Encryption Standard).

Для подальшого розуміння матеріалу одразу введемо кілька необхідних визначень.

Означення 1.1. Відкритий текст - послідовність символів алфавіту, який підлягає шифруванню. Відкритий текст вважається таким, що має сенс, тобто є змістовним.

Означення 1.2. Шифротекстом називається послідовність символів, отримана в результаті застосування до відкритого тексту перетворення зашифрування.

Означення 1.3. Зашифруванням називають перетворення відкритого тексту у шифротекст з використанням ключів.

Означення 1.4. Розшифруванням називають перетворення шифротексту у відкритий текст з використанням таємних ключів.

Означення 1.5. Ключ (в симетричній криптографії) - секретна інформація, що використовується криптографічним алгоритмом для шифрування та/або розшифрування.

Енігма відноситься до роторних машин і складається з двох підсистем - механічної та електричної.

До механічної частини належать клавіатура, набір роторів, розташованих вздовж валу, рефлектора і ступеневого механізму, що обертає один або декілька роторів при натисканні клавіші клавіатури. У воєнної моделі, на відміну від комерційної, також є комутаційна панель,

що замінювала пари букв. Це значно збільшує кількість можливих ключів, а отже і стійкість шифротексту до розшифрування. Електрична частина складається з електричної схеми, що з'єднує клавіатуру, лампочки, ротори та комутаційну панель за її наявності. Сам процес шифрування літер відбувався завдяки електричній частині машини.

У свою чергу кожен ротор являє собою диск зі штирьовими контактами з правої сторони і пласкими контактами з лівої, розташованими по периферії диска, які з'єднані всередині самого ротора. Кожен з контактів відповідає одній з літер алфавіту. При доторку контактів сусідніх роторів замикається електричний ланцюг, струм проходить через утворене коло, внаслідок вмикається лампочка, що відповідає літері шифротексту. Таким чином, кожен ротор виконує шифр моноалфавітної перестановки.

Означення 1.6. Перестановкою довільної множини Σ вважатимемо бієкцію π цієї множини саму у себе, тобто $\pi\colon \Sigma \to \Sigma$.

Приклад 1.1. Нехай дана множина $\Sigma = \{1, 2, 3, 4, 5\}$. Тоді відображення

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{pmatrix}$$

є перестановкою множини Σ.

Означення 1.7. Шифр моноалфавітної заміни - метод симетричного шифрування, при якому кожний символ вхідного повідомлення замінюється на символ шифротексту за допомогою вибраної перестановки на алфавіті, яка є секретним ключем.

Зазвичай в Енігмі використовувалися три або більше ротори і за рахунок їх постійного руху досягали більш складного шифру — поліалфавітної підстановки або поліалфавітної заміни.

Означення 1.8. Поліалфавітна підстановка - сукупність шифрів моноалфавітної перестановки, що можуть бути різними в залежності від

місця букви в тексті.

Загальний принцип роботи шифрувальної машини добре ілюструє рисунок 1.1

Розглянемо детальніше роботи машини.

Нехай при старті роботи оператор натискає клавішу "A". Струм після проходження роторів, рефлектора и знову роторів вмикає лампочку "G". Ця літера стає першою літерою шифротексту. Після цього правий ротор робить оберт, електричний ланцюг змінюється, і при повторному натисканні клавіші "A" струм ввімкне вже іншу лампочку (у даному випадку "C") і так далі.

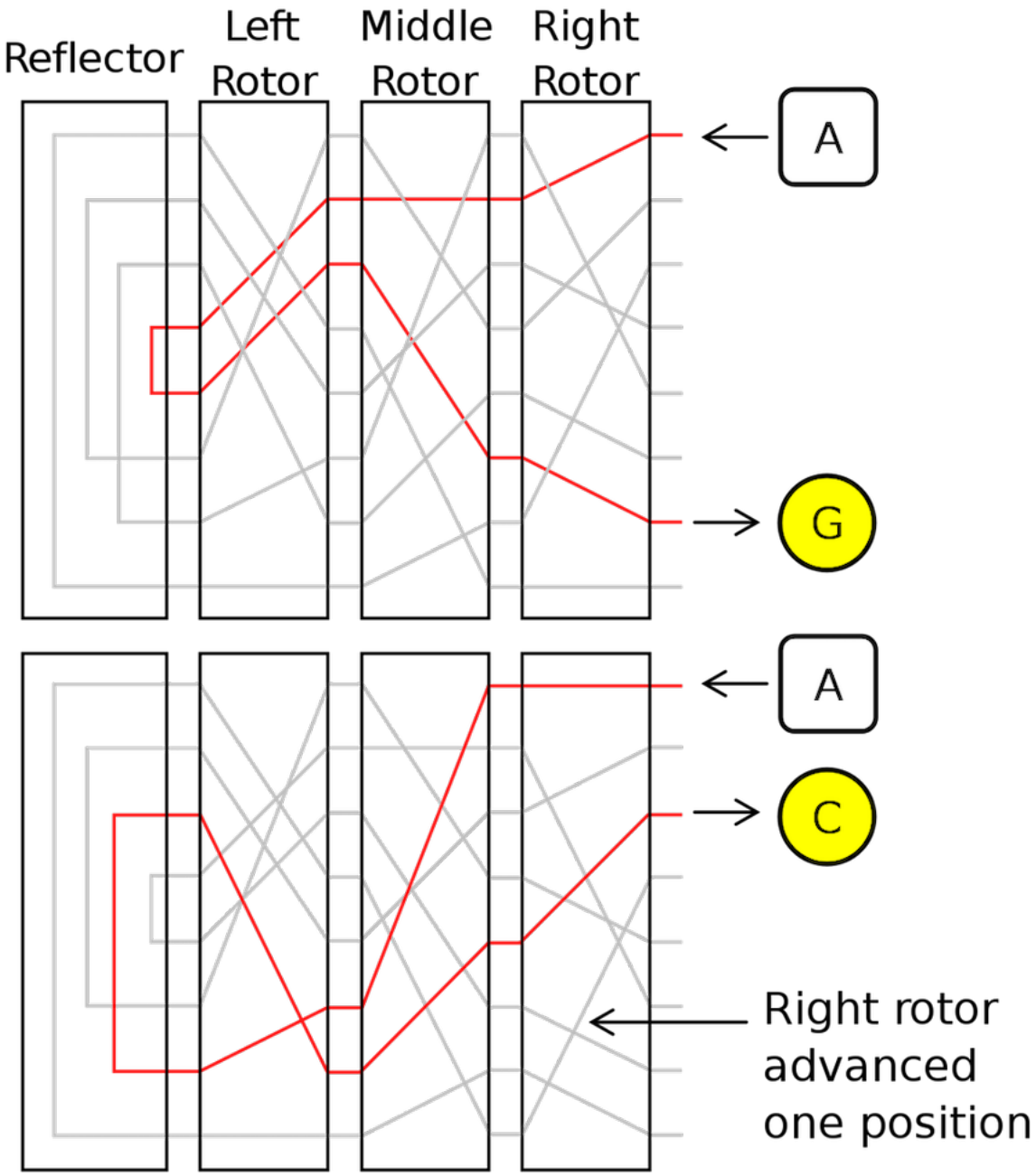
Розглянемо класичну версію Енігми, у якій задіяно три ротори, а алфавіт складається з 26 букв. Це означає, що у кожного роторів по 26 контактів з правої та лівої сторони.

Ключем шифрування для Енігми вважається початкове положення роторів. Оскільки кожен ротор реалізує шифр моноалфавітної перестановки, а порядок розташування роторів може змінюватися, то загальна кількість ключів становить $3! * 26!$. Але не всі з них є хорошими для використання. Наприклад, абсолютно незастосовним є ключ, при якому усі літери замінюються самі на себе (відкритий текст та шифротекст співпадають). Також відомо, що при шифруванні за допомогою Енігми в роторах жодна буква не замінювалася сама у себе, тобто перестановки на алфавітах вибиралися без нерухомих точок, адже це значно спрощує процес розшифровування шифротексту.

Є певний вид перестановок, який якнайкраще підходить для застосування у роторах Енігми для найбільшого ускладнення розшифрування шифротексту. Так як немає загальноприйнятої та загальнопоширеної назви таких перестановок, тут і надалі будемо їх називати перестановками без паралельних перепайок (або перестановками без перепайок). Введемо формальне визначення перестановки без паралельних перепайок.

Означення 1.9. Перестановка $(i_0, i_1, \cdots, i_{n-1})$ множини

Рисунок 1.1 – Принцип роботи Енігми



$\{0, 1, \cdots, n-1\}$ є перестановкою без паралельних перепайок, якщо

$$\forall k, l \in \{0, 1, \cdots, n-1\} : (k + i_k) \pmod n \neq (l + i_l) \pmod n \text{при} k \neq l.$$

Приклад 1.2. Розглянемо множину $\{0, 1, 2\}$ та її перестановку $(1, 2, 0)$. Для них отримаємо такі суми за модулем:

$$(0 + 1) \pmod 3 = 1,$$
$$(1 + 2) \pmod 3 = 0,$$
$$(2 + 0) \pmod 3 = 2.$$

Серед отриманих отриманих сум, побудованих за правилом $(k + i_k) \pmod n$, немає однакових, а отже така перестановка є перестановкою без перепайок.

Нехай послідовність $(t_0, t_1, \cdots, t_{n-1})$ отримано за правилом

$$t_k = k + i_k \pmod n,$$

де $k \in \{0, 1, \cdots, n-1\}$, а $(i_0, i_1, \cdots, i_{n-1})$ - перестановка без перепайок. Зі способу отримання цих сум випливає, що усі $t_k \in \{0, 1, \cdots, n-1\}$ і $t_k \neq t_l$ при $k \neq l$. Фізично це означає, що якщо ротором реалізується перестановка без перепайок, то всередині нього немає жодної пари паралельних з'єднань контактів правої та лівої сторін. Саме тому була обрана відповідна назва.

Чому такі перестановки є бажаними для використання у Енігмі? У випадку, коли ротори не реалізують перестановку без перепайок, то якщо відкритий текст відомий, можна легко отримати рівняння, з яких легше знайти ключ шифрування.

1.2 Основні поняття із теорії імовірності і математичної статистики

Оскільки надалі буде використовуватися один з методів статистичного моделювання для отримання статистичних оцінок характеристик перестановок, а саме метод Монте-Карло, необхідно ввести деякі поняття і терміни. Сам метод буде розглядатися нижче.

Означення 1.10. Сімейство $\mathcal{A}$ підмножин множини X називається сигма-алгеброю, якщо воно задовольняє наступним умовам:

– $\oslash \in \mathcal{A}$;

– $A \in \mathcal{A} \Leftrightarrow \bar{A} \in \mathcal{A}$;

– Об'єднання або перетин зліченної кількості підмножин $\mathcal{A}$ належить до $\mathcal{A}$.

Сигма-алгебри мають важливе значення у теорії імовірностей і математичній статистиці, які будуть використовуватись для оцінки кількості перестановок з певними властивостями.

Також буде використовуватися поняття випадкової величини у деякому імовірнісному просторі.

Означення 1.11. Імовірнісним простором називається трійка $(\Omega, \mathcal{A}, \mathbb{P})$, де Ω - довільна непуста множина, елементи якої називаються елементарними подіями, $\mathcal{A}$ - сигма-алгебра підмножин Ω, $\mathbb{P}$ - імовірнісна міра(або імовірність) така, що $\mathbb{P}(\Omega) = 1$ і $\mathbb{P}(\sum_i A_i) = \sum_i \mathbb{P}(A_i)$, якщо $A_i \bigcap A_j = \oslash$ при $i \neq j$.

Нехай заданий імовірнісний простір $(\Omega, \mathcal{A}, P)$.

Означення 1.12. Випадкова величина - це функція $\xi \colon \Omega \to \mathbb{R}$ така, що

$$\forall x \in \mathbb{R} \quad \{\omega \colon \xi(\omega) < x\} \in \mathcal{A}.$$

Найчастіше на практиці використовуються два різновиди випадкових величин: дискретні та абсолютно неперервні. Дискретна випадкова величина може приймати скінченну або зліченну кількість значень. Вона задається скінченним або зліченним набором значень $\{x_1, x_2, \cdots\}$ та відповідними імовірностями $\{p_1, p_2, \cdots\}$, $p_i = P(\xi = x_i)$, що задовольняють умові $\sum_i p_i = 1$. Якщо ж випадкова величина є абсолютно неперервною, то вона задовольняє умові

$$P(\xi < t) = \int_{-\infty}^{t} p(x)dx,$$

де $p(x)$ - щільність (інтегровна невід'ємна функція, для якої виконується $\int_{-\infty}^{\infty} p(x)dx = 1$).

До числових характеристик випадкових величин належить такі значення, як математичне сподівання, дисперсія та функція розподілу. Вони дають змогу визначити загальні риси закону розподілу випадкової величини.

Нехай заданий імовірнісний простір $(\Omega, \mathcal{A}, P)$ та визначена на ньому випадкова величина ξ.

Означення 1.13. Математичним сподіванням дискретної випадкової величини ξ, що приймає скінченну кількість значень x_i з імовірністю p_i називається сума:

$$M\xi = \sum_i x_i p_i.$$

Математичним сподіванням абсолютно неперервної випадкової величини ξ називається інтеграл:

$$M\xi = \int_{-\infty}^{\infty} xp(x)dx.$$

Математичне сподівання є узагальненим поняттям середнього значення сукупності чисел на той випадок, коли елементи множини значень цієї сукупності мають різну важливість, ціну, пріоритет, вагу тощо, що є характерним для значень випадкової змінної.

Означення 1.14. Дисперсією випадкової величини ξ називається число

$$D\xi = M\xi^2 - (M\xi)^2.$$

Дисперсія показує наскільки сильно значення випадкової величини ξ відрізняються від її математичного сподівання $M\xi$.

У випадку, коли потрібно оцінити характеристики деякої сукупності, але відомі лише її дані, ці дані використовують для оцінки необхідних характеристик.

Означення 1.15. Розподілом випадкової величини називається закон, що описує область значень цієї величини і імовірності їх появи.

Часто розподіл випадкової величини задають за допомогою функції її розподілу.

Означення 1.16. Функція розподілу випадкової величини - це функція

$$F(x) = P(\xi < x)$$

з наступними властивостями:

- $F(x)$ - неспадна функція;
- $\lim_{x\to-\infty} F(x) = 0, \lim_{x\to\infty} F(x) = 1$;
- $F(x)$ є неперервною зліва.

Означення 1.17. Генеральна сукупність - сукупність усіх об'єктів, відносно яких передбачається робити висновки при вивченні певної задачі.

Інакше кажучи, генеральна сукупність складається з усіх об'єктів, які мають якості і властивості, що цікавлять дослідника. Вона може формуватися як за одною, так і за багатьма ознаками.

Означення 1.18. Вибірка $X_1, X_2, \cdots, X_n$ - деяка підмножина генеральної сукупності об'єктів, яка охоплюється експериментом.

Означення 1.19. Статистикою називається будь-яка випадкова величина, що є функцією лише від вибірки $X_1, X_2, \cdots, X_n$.

Вибірки зазвичай характеризуються математичним сподіванням та дисперсією.

На практиці часто параметри розподілів невідомі і доводиться шукати їх приблизні значення для різних невідомих теоретичних характеристик.

Нехай значення випадкової величини утворюють генеральну сукупність, закон розподілу якої є відомим. Однак значення деяких параметрів цього розподілу, такі як математичне очікування або дисперсія, невідомі. Необхідно, вивчаючи вибірки з генеральної сукупності, оцінити невідомий параметр, тобто знайти його приблизне значення.

Означення 1.20. Точковою оцінкою $\hat{\theta}$ деякого параметра θ називається статистика $T(X_1, X_2, \cdots, X_n)$, значення якої при заданій реалізації вибірки $X_1, X_2, \cdots, X_n$ приймають за приблизне значення параметра θ.

Очевидно, що для одного і того самого параметра можна побудувати різні оцінки. Для розуміння якості отриманої оцінки використовують такі властивості, як незміщеність та ефективність.

Означення 1.21. Оцінка $\hat{\theta}$ деякого параметра θ є незміщеною, якщо її математичне сподівання рівне параметру, що оцінюється, тобто $M\hat{\theta} = \theta$.

Серед усіх незміщених оцінок параметра θ найкращою вважається така оцінка θ^*, що має мінімальну можливу дисперсію.

Означення 1.22. Незміщена оцінка θ^* називається ефективною оцінкою, якщо для неї виконується рівність

$$D(\theta^*) = M(\theta^* - \theta)^2 - (M(\theta^* - \theta))^2 \leq D(\hat{\theta}) \forall \hat{\theta}.$$

Чим більший об'єм вибірки, тим точніше значення оцінки визначає невідомий параметр.

Будь-яка точкова оцінка є функцією $T = T(X_1, X_2, \cdots, X_n)$ вибірки $X_1, X_2, \cdots, X_n$, тобто є випадковою величиною, і при кожній реалізації

$x_1, x_2, \cdots, x_n$ вибірки ця функція визначає єдине значення $t = T(x_1, x_2, \cdots, x_n)$ оцінки, яке приймається за приблизне значення параметра. При цьому при кожному експерименті значення оцінки може відрізнятися від дійсного значення параметра. Тому корисно знати і можливу похибку, що виникає при використанні оцінки. Для цього використовують інтервал, у який з високою вірогідністю потрапляє точне значення оцінюваного параметра [1].

Нехай $X_1, X_2, \cdots, X_n$ - вибірка з розподілу $F(\theta)$, де $\theta \in \mathbb{R}$ - невідомий параметр. Також нехай задана $\gamma \in [0, 1]$.

Означення 1.23. Інтервал $[G_1, G_2]$ є довірчим інтервалом з рівнем довіри γ для параметра θ, якщо

$$P(G_1 \leq \theta \leq G_2) = \gamma,$$

де G_1, G_2 - деякі статистики.

Таким чином, довірчий інтервал - це інтервал, побудований за допомогою довільної вибірки з розподілу з невідомим параметром, що містить цей параметр із заданою імовірністю. Зазвичай рівень довіри γ обирають близьким до 1, наприклад 0.9, 0.95, 0.99.

Означення 1.24. Статистичною гіпотезою називають будь-яке твердження про вид або властивості розподілу випадкових величин у рамках експерименту.

Гіпотеза, яка стверджує, що деякий параметр розподілу відповідної сукупності має наперед задане значення або множину значень, називається параметричною гіпотезою.

Якщо для досліджуваного процесу було зформульоване деяке твердження (основна або нульова гіпотеза, гіпотеза H_0), то задача перевірки гіпотези ставиться так: необхідно задати таке правило, що дозволяло б за результатами відповідних експериментів прийняти або відхилити цю гіпотезу [1].

Означення 1.25. Статистичний критерій — це правило, згідно якому щодо гіпотези H_0, що перевіряється, приймається рішення або вважати її не суперечливою статистичним даним або відхилити.

Кожному критерію відповідає деяке розбиття вибіркового простору $\mathcal{A}$ на дві додаткових множини $\mathcal{A}_0$ та $\mathcal{A}_1$ ($\mathcal{A}_0 \bigcap \mathcal{A}_1 = \oslash$, $\mathcal{A}_0 \bigcap \mathcal{A}_1 = \mathcal{A}$), де $\mathcal{A}_0$ - множина точок, для яких H_0 приймається, а $\mathcal{A}_1$- для яких відхиляється. $\mathcal{A}_0$ і $\mathcal{A}_1$ називаються областю прийняття та критичною областю гіпотези H_0 відповідно.

Таким чином, статистичний критерій має вигляд

$$H_0 \text{ відхиляється} \Longleftrightarrow T(X_1, X_2, \cdots, X_n) \in \mathcal{A}_1$$

.

Критична область має бути обрана так, щоб імовірність $P(X_1, X_2, \cdots, X_n \in \mathcal{A}_1 | H_0)$ була малою. Тому при побудові критерію заздалегідь задають деяке мале число α – рівень значущості і накладають умову

$$P(X_1, X_2, \cdots, X_n \in \mathcal{A}_1 | H_0) \leq \alpha.$$

Для перевірки гіпотез про вигляд розподілу випадкової величини ξ використовують критерії згоди. У найпростішому випадку задача ставиться так: нехай $X_1, X_2, \cdots, X_n$ — вибірка з невідомою функцією розподілу $F_\xi(x)$, а гіпотеза H_0: $F_\xi(x) = F(x)$, де $F(x)$ повністю задана. Тоді $H_1 = \overline{H_0}$.

Критерій Колмогорова застосовується при неперервних $F(x)$ та $n \geq 20$. Для пошуку статистики критерію використовують формулу

$$D_n = \sup_{-\infty < x < \infty} |\hat{F}_n(x) - F(x)|,$$

а критерій визначається так

$$H_0 \text{ відхиляється} \Longleftrightarrow D_n \geq t_\alpha,$$

t_α — критична границя для рівня довіри α.

Одним з найбільш універсальних критеріїв є критерій χ^2 Пірсона. Він застосовується з дискретними даними, але оскільки будь-які дані можна звести до дискретних, то цей критерій працює з даними довільної природи [3]. Статистика знаходиться за формулою

$$X_n^2 = \sum_{j=1}^{N} \frac{(\nu_j - np_j)^2}{np_j},$$

де N — кількість інтервалів групування, n — кількість елементів вибірки, $\nu_j = \sum_{i=1}^{n} I(\xi_i = j), p_j = P(\xi = j)$.

Сам критерій такий (за умови, що $n \geq 50$):

$$H_0 \text{ відхиляється} \iff \{X_n^2 > \chi^2_{1-\alpha,N-1}\}.$$

1.3 Випадкові перестановки

Для множин з великою кількістю елементів для генерації усіх перестановок потребуються значні часові та обчислювальні ресурси, що не завжди є раціональними або реалізовними. Наприклад, для множин потужності $n = 16$ загальна кількість перестановок уже становить $n! = 20922789888000$. Зважаючи на те, що в англійському алфавіті 26 літер, а в українському та російському по 33 літери, при нині існуючих технологіях і ресурсах не представляється можливим побудувати та розглянути усі можливі перестановки на даних алфавітах. У таких випадках користуються законом великих чисел.

Твердження 1.1. *У теорії імовірностей закон великих чисел - це принцип, що описує результат виконання одного і того самого експерименту велику кількість разів. Згідно закону, середнє значення*

скінченної вибірки $(X_1, X_2, \cdots, X_n)$ *з фіксованого розподілу близьке до математичного очікування цього розподілу.*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to MX \text{ при } n \to \infty.$$

Закон застосовний тільки тоді, коли розглядається достатньо велика серія експериментів.

Це означає, що при неможливості розглянути генеральну сукупність, можна зформувати вибірку з цієї сукупності, а результат дослідження отриманої вибірки узагальнити на генеральну сукупність. Очевидно, що чим більший об'єм вибірки, тим більш репрезентативною вона виявиться за рахунок зменшення впливу аномальних значень окремих елементів(так званих викидів). Також важливою вимогою до формування вибірки є те, що елементи з генеральної сукупності для вибірки обираються випадковим чином.

Для дослідження перестановок із заданими характеристиками і оцінки кількості таких перестановок використовуються випадкові перестановки, які є елементами сукупності усіх перестановок множини.

Означення 1.26. Нехай є множина $\{0, 1, \cdots, n-1\}$. Випадковою перестановкою даної множини називається вектор $(\xi_0, \xi_1, \cdots, \xi_{n-1})$, усі елементи якого приймають значення від 0 до $n-1$, при цьому імовірність збігу двох довільних елементів дорівнює 0, а ймовірність кожної перестановки — $\frac{1}{n!}$.

Як приклад випадкової перестановки можна навести результат перетасовування колоди карт.

Використання випадкових перестановок часто є базою у таких областях, як криптографія, теорія кодування або моделювання, де використовуються імовірнісні алгоритми.

Якщо об'єм вибірки випадкових перестановок достатній, то, за законом великих чисел, можна вважати, що частка перестановок із

заданою характеристикою у вибірці співпадає із їх часткою серед усіх можливих перестановок. Для генерації випадкових перестановок у рамках даної роботи використовується алгоритм Фішера-Йетса [2], який реалізує випадкове перетасовування елементів множини. Він є незміщеним, а отже кожна перестановка генерується з однаковою імовірністю, і забезпечує ефективне і швидке отримання випадкових перестановок. Алгоритм не потребує додаткової пам'яті, а час його роботи пропорційний розміру множини, тож при коректній реалізації він вважається оптимальним для використання. Сам алгоритм виглядає таким чином:

Вхід: множина $\{0, 1, \cdots, n-1\}$.

1. Для усіх i від $n-1$ до 1 виконати:

1.1 j - випадкове число від 0 до i

1.2 Поміняти місцями i-тий та j-тий елементи

Вихід: випадкова перестановка.

На кожній ітерації обирається випадковий елемент із усіх залишившихся, тобто на першій ітерації є n способів обрати елемент, на другій $n-1$ способів обрати другий елемент і так далі до останньої ітерації, на якій є два способи обрати останній елемент. На жодній ітерації немає можливості обрати вже обраний раніше елемент, оскільки усі вони переносяться у кінець масиву шляхом перестановки з останнім необраним елементом. Послідовність довжини n можна отримати $n \times (n-1) \times (n-2) \times \cdots \times 1 = n!$ способами, що співпадає із загальною кількістю перестановок. Це значить, що імовірність отримати довільну перестановку дорівнює $\frac{1}{n!}$, тож усі перестановки рівноімовірні.

Як і усі інші випадкові процеси, алгоритм Фішера-Йетса залежить від використовуваного генератора випадкових або псевдовипадкових чисел. Генерація псевдовипадкової послідовності повністю задається початковим станом генератора при старті роботи. Генератор псевдовипадкових чисел не може створити більше перестановок, ніж число внутрішніх станів генератора. Навіть коли число можливих станів перевищує число перестановок, деякі перестановки можуть з'являтися

частіше за інші. Для запобігання появи нерівномірності розподілу кількість внутрішніх станів генератора має бути на кілька порядків більша кількості перестановок. Зазвичай, мови програмування і бібліотеки використовують 32-бітне число для внутрішніх станів, що означає, що генератор може створити 2^{32} різноманітних випадкових чисел. Тож у при генерації випадкових перестановок це зауваження має бути враховано для покращення якості вибірки.

1.4 Метод Монте-Карло

Як вже зазначалося раніше, для того, щоб отримати статистичні оцінки характеристик перестановок, буде використовуватися метод Монте-Карло — чисельний метод для вивчення випадкових процесів. Він полягає у багаторазовому моделюванні за допомогою генератора випадкових величин необхідного процесу, і на основі згенерованих даних отримуються імовірнісні характеристики вирішуваної задачі. Значною перевагою даного методу є те, що в його основі лежить закон великих чисел (середнє значення скінченної вибірки із фіксованого розподілу прямує до математичного сподівання цього розподілу). Це дозволяє врахувати у моделі процесу елемент випадковості і складність реального світу.

Основними кроками методу є:

1) Визначення області можливих даних.

2) Генерація випадкових вхідних даних із області можливих даних використовуючи деякий заданий розподіл імовірностей.

3) Виконання обробки вхідних даних, вираховування імовірнісних характеристик вирішуваної задачі.

4) Отримання кінцевого результату із проміжних розрахунків.

Наведемо приклад застосування методу Монте-Карло.

Приклад 1.3. Нехай необхідно визначити середню відстань між двома випадковими точками у колі за допомогою методу Монте-Карло. Для цього потрібно виконати такі дії:

– взяти випадкову пару точок,

– вирахувати для неї відстань між точками.

При достатньо великій кількості пар після усереднення отриманих відстаней знайдемо шукану середню відстань.

Для ефективного застосовування методу Монте-Карло важливо враховувати два моменти. По-перше, якщо випадкові величини не є незалежними, наближення процесу буде мінімальним. По-друге, вхідних даних має бути досить багато. Це дозволяє зменшити вплив викидів (значень, що виділяються із загальної вибірки) на кінцевий результат настільки, що ним можна знехтувати. При дотриманні цих умов можна отримати найбільш точну модель процесу.

Метод Монте-Карло є простим, ефективним і може використовуватися при будь-якому розподілі, проте для його застосування потрібен хороший генератор випадкових чисел, інакше вибірка може виявитися нерепрезентативною. У такому випадку кінцеві результати не можна узагальнювати на генеральну сукупність, з якої була обрана вибірка. Ще одним недоліком є проблема визначення об'єму вибірки, тобто кількості точок для вирішення задачі із заданою точністю. Зріст кількості згенерованих даних може призвести до значного ускладнення обчислень і обробки цих даних. Необхідний баланс між точністю та складністю обчислень обирається в залежності від постановки задачі дослідження і наявних ресурсів.

У рамках дослідження класів характеристик перестановок метод буде застосовуватися наступним чином.

1) Задамо вхідну множину перестановок степені n M і її потужність $n! = |M|$.

2) Визначимо характеристики перестановок та розбиття на класи.

3) Генеруватимемо випадкову перестановку множини M і шукатимемо її характеристику за наведеним далі алгоритмом. Для кожної характеристики статистично з допомогою метода Монте-Карло рахуватимемо кількість перестановок, з такою характеристикою.

4) Виділимо класи з різними характеристиками, опишемо властивості перестановок кожного класу та їх застосовність у роторних шифраторах.

5) Представимо кінцеві результати дослідження.

1.5 Оцінка кількості перестановок без паралельних перепайок

Серед нечисленних робіт, присвячених оцінці кількості перестановок без паралельних перепайок, відзначимо такі статті, як [3, 4, 5, 6], у яких автори називають такі перестановки повними відображеннями, повними перестановками або "хорошими" перестановками.

Нехай задана множина $\Sigma = \{0, 1, \cdots, n-1\}$, яка містить n елементів.

У теоремі 1 в статті [4] зауважується, що якщо n парне, то кількість перестановок без перепайок дорівнює нулю, тому зазвичай множини з парною кількістю елементів не беруть до уваги при дослідженні таких перестановок. Значно більше зацікавленості викликають множини, у яких n - непарне. Також у [4] отримано верхню границю кількості перестановок без паралельних перепайок: вірогідність, що випадкова перестановка виявиться перестановкою без перепайок не перевищує $P(n) = e^{-cn}$ при достатньо великих непарних n, де $c \geq 0.08854$. Таке значення було отримано комбінаторним методом. У той же час нерівність із теорії стохастичних процесів дає більш слабку оцінку — $c \geq 0.06766$. Автори відзначили, що якщо розглядати не усі можливі перестановки, а

лише ті, які можна вважати випадковими, то матимемо асимптотичну границю $e^{-(1-\varepsilon)n}$, де ε - додатне довільно мале число.

У статті [5] цю оцінку величини c було покращено. Автор довів, що $c \geq \frac{\ln 2}{2} \approx 0.35$. Це означає, що загальна кількість перестановок без паралельних перепайок не перевищує $(n+1)! \cdot 2^{-m+O(1/\sqrt{\ln m})}$, де $m = \frac{n-1}{2}$ - ціле число. Також автори навели алгоритм перечислення усіх перестановок без перепайок, що значно полегшує пошук таких перестановок.

У [6] вводиться таке поняття, як k-good permutation(k-хороша перестановка).

Означення 1.27. Нехай $s(k) = (s_0, s_1, \cdots, s_{k-1})$ - часткове представлення перестановки довжини n. Будемо вважати це представлення k-хорошим, якщо значення $t_l(l + s_l) \pmod n$ усі різні при різних $l \leq k$.

Очевидно, що перестановка не може бути перестановкою без перепайок, якщо усі її часткові представлення не k-хороші для усіх k. Приводяться такі результати. Для $n = 25$ кількість 10-хороших перестановок починаючи з нуля була знайдена перебором і дорівнює $58,460,060,880$. Виходячи з цього, вірогідність, що загальна перестановка довжини 25 є перестановкою без перепайок оцінюється щонайменше числом $1.86785 \cdot 10^{-9}$ [6]. Позначивши $\pi(n) = e^{-c_n n}$, автори отримують деяку оцінку для c_n, $n = 25$. Відповідне значення для $n = 25$ оцінюється як $c_{25} \leq 0.8039$. Приводиться і більш точне значення, базоване на середній кількості перестановок без перепайок на деяких проміжках, — $\hat{c_{25}} = 0.789$. У статті приводиться таблиця точних значень P_n та $c_n = -\frac{1}{n} \log P(n)$ для непарних n від 1 до 19 ([6, табл. 2]).

Автори роблять висновок, що для великих n справедливе рівняння $P(n) = e^{-c_n/n}$.

Також були приведені оцінки P_n та c_n для $n = 25, 35, 45, 55$ та ∞ за умови, що $P_n = e^{-c_n/n}$ ([6, табл. 3]). У [6] приведений аналіз перестановок без перепайок для застосування у криптографії, зокрема у роторних

шифрувальних машинах.

У [7] для оцінки кількості перестановок без перепайок пропонується використовувати метод пришвидшеного моделювання. Оскільки сам алгоритм отримання перестановок та його приведена модифікація у рамках даного дослідження не є принциповими, тут він наводитися не буде, але з ним можна ознайомитися у розділі "Алгоритм ускоренного моделирования" [7]. Не зважаючи на те, що метод є досить простим для реалізації, він дозволяє при відносно невеликих витратах часу побудувати незміщені оцінки і відповідні довірчі інтервали для P_n при досить великих n (у статті [7] приводяться оцінки для P_n до $n = 155$ включно). Більш того, даний алгоритм дозволяє у практичних підрахунках підтвердити співвідношення $P_n \sim ae^{-cn}$ і вказати більш точні границі для c, а саме у [7] приводяться такі: $0.9825 \leq c \leq 0.9883$.

Загалом, для $n \geq 75$ для P_n були знайдені та указані наступні межі:

$$413.099exp\{-0.9883n\} \leq P_n \leq 267.384exp\{-0.9825n\},$$

а наведені чисельні дані для деяких n свідчать про високу степінь точності нижніх та верхніх оцінок і можливість з їх допомогою прогнозувати значення P_n.

Висновки до розділу 1

У розділі описано будову і принцип роботи шифрувальної машини Енігма, ідею алгоритму шифрування, що у ній використовується та утворювані нею поліалфавітні перестановки. Введено поняття перестановки без перепайок, їх застосування в Енігмі, оглянуто існуючі оцінки кількості перестановок без перепайок серед усіх можливих перестановок на деякій множині. Крім того, розглянутий метод

Монте-Карло статистичного моделювання і зауваження до його використання.

2 КЛАСИ ПЕРЕСТАНОВОК ЗІ СПЕЦІАЛЬНИМИ ВЛАСТИВОСТЯМИ ТА ОЦІНЮВАННЯ ЇХ ПОТУЖНОСТІ

У даному розділі розглядаються довірчі інтервали для параметрів біноміального розподілу, нормального розподілу та розподілу Пуассона, розглядаються поняття характеристики перестановки, вводиться класифікація перестановок за їх характеристиками, приводиться точна кількість перестановок кожного класу для множин з потужністю від 3 до 11 включно і порівнюється отримана кількість перестановок без перепайок із вже відомими даними для таких множин. Також застосовано метод статистичного моделювання для отримання оцінок кількості перестановок у класах для множин з більшою потужністю. Виконано порівняння різних статистичних апроксимацій для побудови довірчих інтервалів для потужностей класів. Проведено аналіз результатів.

2.1 Довірче оцінювання

Задача побудови довірчого інтервалу для деякого параметра θ розподілу і задача перевірки гіпотези відносно цього параметра еквівалентні.

Оберемо довільне $\gamma \in [0, 1]$.

Розглянемо нормальний розподіл $\mathcal{N}(a, \sigma^2)$.

Нехай необхідно побудувати довірчий інтервал для середнього вибірки $X_1, X_2, \ldots, X_n$ з розподілу $\mathcal{N}(a, \sigma^2)$ при відомому σ^2. Відомо, що

$$\sqrt{n}\frac{\overline{X} - a}{\sigma} \sim \mathcal{N}(0, 1),$$

де $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ([8]). Нехай z_q — q-квантиль стандартного нормального розподілу. Тоді в силу його симетрії

$$P(-z_{1-\frac{\gamma}{2}} < \sqrt{n}\frac{\overline{X}-a}{\sigma} < z_{1-\frac{\gamma}{2}}) = 1-\gamma \Longrightarrow$$
$$\Longrightarrow P(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\gamma}{2}} < a < \overline{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\gamma}{2}}) = 1-\gamma, \quad (2.1)$$

звідки довірчий інтервал такий:

$$(\overline{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\gamma}{2}}; \overline{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\gamma}{2}}).$$

У тому випадку, коли дисперсія σ^2 невідома, випадкова величина $\sqrt{n}\frac{\overline{X}-a}{s}$ розподілена за законом Стьюдента з $n-1$ ступенями свободи, де $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$ ([8]). Тоді якщо $t_{q,n-1}$ — квантиль розподілу Стьюдента, то в силу його симетрії

$$P(-t_{1-\frac{\gamma}{2},n-1} < \sqrt{n}\frac{\overline{X}-a}{s} < t_{1-\frac{\gamma}{2},n-1}) = 1-\gamma \Longrightarrow$$
$$\Longrightarrow P(\overline{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\gamma}{2},n-1} < a < \overline{X} + \frac{s}{\sqrt{n}} t_{1-\frac{\gamma}{2},n-1}) = 1-\gamma, \quad (2.2)$$

а довірчий інтервал буде таким:

$$(\overline{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\gamma}{2},n-1}; \overline{X} + \frac{s}{\sqrt{n}} t_{1-\frac{\gamma}{2},n-1})$$

.

Якщо ж необхідно побудувати довірчий інтервал для дисперсії σ^2 за умови, що середнє відоме, то користуючись тим, що випадкова величина $\frac{\sum_{i=1}^{n}(X_i-a)^2}{\sigma^2}$ має розподіл $\chi^2(n)$ та знайшовши γ-квантиль цього розподілу $\chi^2_{\gamma,n}$, при обраному γ отримуємо

$$P(\chi^2_{\frac{1-\gamma}{2},n} \le \frac{\sum_{i=1}^{n}(X_i-a)^2}{\sigma^2} \le \chi^2_{\frac{1+\gamma}{2},n}) = \gamma.$$

Після нескладних перетворень:

$$P(\frac{\sum_{i=1}^{n}(X_i-a)^2}{\chi^2_{\frac{1+\gamma}{2},n}} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(X_i-a)^2}{\chi^2_{\frac{1-\gamma}{2},n}}) = \gamma$$

. Тоді довірчий інтервал такий ([8]):

$$(\frac{\sum_{i=1}^{n}(X_i-a)^2}{\chi^2_{\frac{1+\gamma}{2},n}}; \frac{\sum_{i=1}^{n}(X_i-a)^2}{\chi^2_{\frac{1-\gamma}{2},n}}).$$

Для біноміального розподілу $Bin(n,p)$ та розподілу Пуассона $Pois(\lambda)$, використовуючи розподіл точкової оцінки параметрів, отримуємо відповідні довірчі інтервали ([8]):

$$\overline{X} \pm \frac{c_\gamma}{\sqrt{n}}\sqrt{\overline{X}(1-\overline{X})}, \quad \overline{X} \pm c_\gamma\sqrt{\frac{\overline{X}}{n}},$$

де $c_\gamma = \Phi^{-1}\left(\frac{1+\gamma}{2}\right)$.

2.2 Перехід від біноміального розподілу до інших розподілів

Оскільки при генерації перестановок за методом Монте-Карло кожна перестановка з'являється з однаковою імовірністю, то число появ певної характеристики перестановок у послідовності незалежних генерацій має біноміальний розподіл, а оцінкою параметра p буде $\frac{n}{N}$, де n — кількість появ характеристики, N — кількість генерацій перестановок. Функція розподілу має вигляд

$$F(x,N,p) = \sum_{i=1}^{x} C_N^i p^i (1-p)^{N-i}$$

— імовірність появи характеристики менше або рівне x разів серед N випробувань ([9]). Середнє та дисперсія такого розподілу:

$$M(x) = Np, D(x) = Np(1-p).$$

При $N \to \infty, p \to 0, Np = const$ біноміальний розподіл можна звести до розподілу Пуассона із параметом $\lambda = Np$ ([9]). Оскільки розподіл Пуассона пов'язаний з розподілом χ^2, то функція біноміального розподілу може бути представлена у вигляді

$$F(x, N, p) = \sum_{i=1}^{x} \frac{(Np)^i}{i!} e^{-Np} = P_{\chi^2}(2Np, 2x+2),$$

де $P_{\chi^2}(u, k)$ — функція розподілу χ^2 з k степенями свободи.

При $N \to \infty$ біноміальний розподіл зводиться до нормального із середнім $a = Np$ та дисперсією $\sigma^2 = Np(1-p)$. Такий перехід задовільний при $Np(1-p) > 5$ і $0.1 \leq p \leq 0.9$ або при $Np(1-p) > 25$ для довільного p ([9]). Таким чином, функція розподілу

$$F(x, N, p) = \Phi\left(\frac{x - Np + 0.5}{\sqrt{Np(1-p)}}\right),$$

де $\Phi(u)$ — функція стандартного нормального розподілу ([9]).

2.3 Характеристики перестановок, їх побудова, властивості та критерії їх класифікації

До цих пір перестановки поділялися лише на перестановки з перепайками та перестановки без перепайок, при цьому мова велася про те, що останні є найбажанішими до використання у криптографії. Проте і серед перестановок з перепайками є такі, що досить успішно можуть бути

застосовані наприклад у роторних шифраторах, зокрема в Енігмі. Степінь застосовності перестановки для реалізації шифру залежить від характеристики даної перестановки.

Нехай є множина $\{0, 1, \ldots, n-1\}$ і деяка її перестановка $(i_0, i_1, \ldots, i_{n-1})$. За правилом

$$\begin{array}{cccccc} & 0 & 1 & \ldots & n-1 \\ \boxplus & i_0 & i_1 & \ldots & i_{n-1} \\ \hline & j_0 & j_1 & \ldots & j_{n-1} \end{array}$$

отримуємо послідовність $(j_0, j_1, \ldots, j_{n-1})$ залишків суми за модулем n. Усі $j_t, t = \overline{0, n-1}$ належать множині $\{0, 1, \ldots, n-1\}$ і не обов'язково є різними. Нехай k_0 — кількість нулів серед $(j_0, j_1, \ldots, j_{n-1})$, k_1 — кількість одиниць, …, k_{n-1} — кількість $n-1$. Інакше кажучи, перепишемо множину $\{j_0, j_1, \ldots, j_{n-1}\}$ у вигляді мультимножини:

$$\{0^{k_0}, 1^{k_1}, \ldots, (n-1)^{k_{n-1}}\}.$$

Нескладно бачити, що для $\{k_0, k_1, \ldots, k_{n-1}\}$ має виконуватися

$$\sum_{j=0}^{n-1} k_j = n.$$

Означення 2.1. Характеристикою перестановки $(i_0, i_1, \ldots, i_{n-1})$ множини $\{0, 1, \ldots, n-1\}$ з відповідним набором $\{k_0, k_1, \ldots, k_{n-1}\}$ послідовністю будемо називати послідовність $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$, де α_i — число чисел k_j, що трапляються i разів.

Для характеристик перестановок справедлива наступна рівність:

$$\sum_{i=0}^{n+1} i\alpha_i = n.$$

Наведемо приклад побудови характеристики перестановки.

Приклад 2.1. Нехай є множина $\{0, 1, 2, 3, 4, 5, 6\}$.

Візьмемо довільну її перестановку, наприклад $(3, 1, 4, 2, 5, 0, 6)$.

Знайдемо послідовність залишків сум за модулем 7:

$\boxplus$	0	1	2	3	4	5	6
	3	1	4	2	5	0	6
	3	2	6	5	2	5	5

Представимо отриману послідовність у вигляді мультимножини: $\{0^0, 1^0, 2^2, 3^1, 4^0, 5^3, 6^1\}$. Нескладно бачити, що рівність $\sum_{j=0}^{n-1} k_j = n$ виконується $(0 + 0 + 2 + 1 + 0 + 3 + 1 = 7)$.

Перейдемо до побудови характеристики.

Нуль разів зустрічаються 3 числа — 0, 1 та 4, тому $\alpha_0 = 3$. Один раз зустрічаються числа 3 та 6 $(\alpha_1 = 2)$. Двічі зустрічається тілки число 2 $(\alpha_2 = 1)$. Тричі зустрічається число 5 $(\alpha_3 = 1)$. Чотири, п'ять та шість разів не зустрічається жодне з чисел, тому $\alpha_4 = \alpha_5 = \alpha_6 = 0$.

Таким чином, характеристикою перестановки $(3, 1, 4, 2, 5, 0, 6)$ множини $\{0, 1, 2, 3, 4, 5, 6\}$ є послідовність $(3, 2, 1, 1, 0, 0, 0)$. Перевіримо виконання рівності $\sum_{i=0}^{n-1} i\alpha_i = n$:

$$0 \cdot 3 + 1 \cdot 2 + 2 \cdot 1 + 3 \cdot 1 + 4 \cdot 0 + 5 \cdot 0 + 6 \cdot 0 = 7.$$

Рівність вірна. Характеристика побудована.

Характеристики кількох перестановок можуть співпадати. Це означає, що ці перестановки мають однакові властивості, а отже криптоаналіз систем, побудованих на них, має складність одного порядку.

Очевидно, що чим більше значення другої позиції характеристики перестановки (тобто чим більше унікальних лишків суми за модулем n), тим краще ця перестановка для використання у якості підключа шифрування у роторних шифраторах.

2.4 Клас перестановок без перепайок

Так як для перестановки без перепайок усі залишки $j_t, t = \overline{0, n-1}$ є унікальними і трапляються лише один раз, то $\alpha_1 = n$, а характеристика такої перестановки — це послідовність виду $\{0, n, 0, \ldots, 0\}$

Про цей клас перестановок вже відомо досить багато. Частина відомостей була розглянута у першому розділі, в тому числі і статистичні оцінки кількості перестановок даного класу для множин перестановок різних потужностей. Порівняємо ці оцінки з точною кількістю перестановок без перепайок для малих n.

Алгоритм підрахунку є таким (1).

Algorithm 1 Алгоритм підрахунку кількості перестановок без перепайок

1: Встановлюємо лічильник $c = 0$

2: Будуємо множину $\{0, 1, \ldots, n-1\}$ та множину усіх можливих її перестановок

$$I = \{(i_0, i_1, \ldots, i_{n-1}) \colon (i_0, i_1, \ldots, i_{n-1}) \text{ — перестановка множини } \{0, 1, \ldots, n-1\}\}$$

3: Для кожної перестановки $(i_0, i_1, \ldots, i_{n-1}) \in I$ отримуємо послідовність $(j_0, j_1, \ldots, j_{n-1})$ таку, що

4: Якщо $\forall k, l \in \{0, 1, \ldots, n-1\}, k \neq l \colon j_k \neq j_l$, збільшуємо лічильник c на 1.

Очевидно, що, знаючи потужність класу перестановок без перепайок, легко знайти імовірність того, що випадкова перестановка виявиться перестановкою без перепайок. Для цього достатньо розділити потужність класу на загальну кількість перестановок:

$$P_n = \frac{|\text{клас перестановок без перепайок}|}{n!}.$$

За допомогою наведеного вище алгоритму були отримані результати для $n = 3, 5, 7, 9, 11$, наведені у таблиці 2.1 (з точністю до 9 знаків після крапки).

Таблиця 2.1 – Точна і змодельована кількість перестановок без перепайок

n	Загальна кількість перестановок	Кількість перестановок без перепайок	Точне P_n	Оцінка P_n
3	6	3	0.5	0.5
5	120	15	0.125	0.125
7	5040	133	0.0263889	0.0263889
9	362880	2025	0.00558036	0.00558036
11	39916800	37851	0.000948247	0.000948247

Як можна бачити, оцінки імовірностей появи перестановки без перепайок, отримані у [6], повністю співпадають з отриманими у рамках цієї роботи точними значеннями цих імовірностей для $n = \overline{3, 11}$.

Для перестановок довжини $n = 13 - 19, 25, 35, 45, 55$ також можна отримати приблизну імовірність появи перестановки без перепайок за допомогою методу Монте-Карло. Він дасть змогу оцінити частку перестановок необхідного класу серед усіх перестановок множини за умови великої кількості експериментів та закону великих чисел.

Задамо послідовність дій.

Algorithm 2 Алгоритм підрахунку кількості перестановок без перепайок для великих n методом Монте-Карло

1: Встановлюємо кількість експериментів N — велике число

2: Встановлюємо лічильник $c = 0$

3: Будуємо множину $\{0, 1, \ldots, n-1\}$

4: Будуємо випадкову незалежну перестановку $(i_0, i_1, \ldots, i_{n-1})$ множини $\{0, 1, \ldots, n-1\}$

5: Для перестановки $(i_0, i_1, \ldots, i_{n-1}) \in I$ отримуємо послідовність $(j_0, j_1, \ldots, j_{n-1})$ таку, що

$$\begin{array}{cccccc} & 0 & 1 & \ldots & n-1 \\ \boxplus & & & & \\ & i_0 & i_1 & \ldots & i_{n-1} \\ \hline & j_0 & j_1 & \ldots & j_{n-1} \end{array}$$

6: Якщо $\forall k, l \in \{0, 1, \ldots, n-1\}, k \neq l$: $j_k \neq j_l$, збільшуємо лічильник c на 1

7: Зменшуємо N на 1

8: Повторюємо кроки 3) - 6) доти, поки $N \geq 0$

9: Визначаємо імовірність появи перестановки без перепайок за формулою

$$P_n = \frac{c}{\text{кількість експериментів}} = \frac{c}{N}$$

Алгоритм 2 із встановленою кількістю експериментів $N = 10^7$ був використаний для побудови порівняльної таблиці імовірностей появи перестановки без перепайок множин довжини $n = 13 - 19, 25, 35, 45, 55$. Наведемо ці результати в порівнянні з результатами, отриманими у [6] (2.2).

Таблиця 2.2 – Порівняльна таблиця результатів

n	Отримана оцінка P_n	Відома оцінка P_n
11	0.0009506	0.000948247
13	0.0002991	0.000165467
15	0.0000498	0.0000278096
17	0.0000052	0.00000451522
19	0.0000016	0.000000720595
25	0	2.73×10^{-9}
35	0	2.25×10^{-13}
45	0	1.75×10^{-17}
55	0	1.32×10^{-21}

У даному випадку імовірності вже не збігаються цілком, проте порядок величин є однаковим для $n = \overline{13, 19}$, а для $n = 25, 35, 45, 55$ відома оцінка дуже близька до нуля. Різницю в приведених результатах можна пояснити різною кількістю згенерованих перестановок та імовірнісною природою появи перестановки у даному алгоритмі. У випадку $n = 25, 35, 45, 55$ отримані оцінки імовірностей пояснюються недостатньою кількістю експериментів для того, щоб хоча б одна згенерована перестановка виявилася представником класу перестановок без перепайок, адже потужність класу є досить малою.

2.5 Клас перестановок з виродженою характеристикою

Означення 2.2. Престановкою множини довжини n з виродженою характеристикою будемо називати таку перестановку, характеристика якої має вигляд $\{n-1, 0, \ldots, 0, 1\}$.

Фактично це означає, що усі лишки суми за модулем n є однаковими, що значно спрощує криптоаналіз шифру. У випадку відомого шифротексту та частини ключа або відкритого тексту пошук повного повідомлення полягає у вирішенні системи рівнянь виду

$$i + j_i \pmod{n} = k,$$

де i — символ повідомлення, j_i — символ ключа (перестановки), а k — лишок суми за модулем, який є однаковим для усіх рівнянь системи.

Частка перестановок із виродженою характеристикою серед усіх можливих перестановок множини дуже мала. В результаті проведених експериментів для $n = 13 - 19, 25, 35, 45, 55$ найбільша їх кількість була отримана при $n = 11$ ($P_{11} = 0.0000014$), При інших n зустрічалася лише одна або не зустрічалися зовсім.

На відміну від перестановок без перепайок, які є найбільш вдалими для застосування у роторних шифраторах, перестановки з виродженою характеристикою є зовсім не застосовними.

2.6 Побудова довірчих інтервалів для потужностей класів перестановок різної довжини

Окрім перестановок «без перепайок» та перестановок з виродженою характеристикою були виділені інші класи перестановок. При $n = 11$ можливо здійснити повний перебір перестановок, тому можна отримати точні значення потужності усіх класів. Для цього використовувався наступний алгоритм (3):

Algorithm 3 Алгоритм побудови класів методом повного перебору перестановок

1: Будуємо множину $\{0, 1, \ldots, n-1\}$

2: Будуємо множину усіх перестановок множини $\{0, 1, \ldots, n-1\}$:

$$I = \{(i_0, i_1, \ldots, i_{n-1}) \text{ — перестановка множини } \{0, 1, \ldots, n-1\}\}$$

3: Для кожної перестановки $(i_0, i_1, \ldots, i_{n-1}) \in I$:

4: Отримуємо послідовність $(j_0, j_1, \ldots, j_{n-1})$ таку, що

5: З послідовності $(j_0, j_1, \ldots, j_{n-1})$ знаходимо характеристику $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$

6: Якщо характеристика $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ не зустрічалася на попередніх ітераціях, встановлюємо лічильних $c = 1$. Якщо зустрічалася — збільшуємо c на 1

У випадку, коли повністю перебрати перестановки неможливо і знайти точні потужності класів не вдасться, при їх оцінці необхідно побудувати довірчий інтервал для потужності кожного з класів, а отже, знайти їх точкові оцінки. Для цього потрібно виконати такі кроки:

Algorithm 4 Алгоритм побудови класів методом Монте-Карло

1: Встановлюємо кількість експериментів N — велике число

2: Будуємо множину $\{0, 1, \ldots, n-1\}$

3: Поки $N \geq 0$:

4: Будуємо випадкову незалежну перестановку $(i_0, i_1, \ldots, i_{n-1})$ множини

5: Для перестановки $(i_0, i_1, \ldots, i_{n-1})$ отримуємо послідовність $(j_0, j_1, \ldots, j_{n-1})$ таку, що

$$\begin{array}{cccccc} & 0 & 1 & \ldots & n-1 \\ \boxplus & & & & \\ & i_0 & i_1 & \ldots & i_{n-1} \\ \hline & j_0 & j_1 & \ldots & j_{n-1} \end{array}$$

6: З послідовності $(j_0, j_1, \ldots, j_{n-1})$ знаходимо характеристику $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$

7: Якщо характеристика $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ не зустрічалася на попередніх ітераціях, встановлюємо лічильних $c = 1$. Якщо зустрічалася — збільшуємо c на 1

8: Зменшуємо N на 1

Беручи до уваги закон великих чисел

$$\frac{1}{N} \sum_{i=1}^{N} X_i \to MX \text{ при } N \to \infty,$$

можна вважати, що отримані кількості появ характеристик являються середнім значенням для кожної характеристики. За допомогою точкових оцінок, отриманих за описаним вище алгоритмом легко отримати довірчі інтервали для кожного з класів.

Нехай N — кількість експериментів (генерацій випадкових перестановок), k — точкова оцінка потужності певного класу (кількість перестановок з характеристикою $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$). Тоді при випадковому виборі перестановки імовірність, що $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ виявиться її характеристикою дорівнює $p = \frac{k}{N}$, а точкова оцінка потужності класу — $n! \cdot p$.

У багатьох мовах програмування існують реалізовані функції для

побудови довірчих інтервалів різних розподілів, в тому числі і біноміального розподілу. В рамках даного дослідження використовувався пакет scipy із мови програмування Python, і довірчий інтервал був отриманий за допомогою функції $stats.binom.interval(1-\alpha, N, p)$, де α — рівень значущості.

Також побудований довірчий інтервал для параметра p в біноміальній моделі при використанні наступних формул:

$$p_{\text{нижнє}} = \frac{k}{NR_1} \tag{2.3}$$

$$p_{\text{верхнє}} = \frac{k}{NR_2}, \tag{2.4}$$

де, як вказано в [11],

$$R_1 = \frac{k(2N - k + 1 + \frac{1}{2}\chi_\alpha)}{N\chi_\alpha},$$

$$R_2 = \frac{k(2N - k + \frac{1}{2}\chi_{1-\alpha})}{N\chi_{1-\alpha}}.$$

Помноживши імовірності (2.3) і (2.4) на загальну кількість перестановок довжини 11, отримаємо довірчі інтервали для потужностей класів.

У таблиці 2.3 приведені результати побудови довірчих інтервалів для потужності усіх класів в біноміальній моделі з використанням засобів мови програмування Python та формул (2.3) і (2.4).

Також для отримання довірчих інтервалів використовувалися апроксимації біноміального розподілу до нормального розподілу та розподілу Пуассона з врахуванням обмежень, описаних на початку глави.

Таблиця 1 – Довірчі інтервали для потужності усіх класів перестановок довжини $n = 11$ в моделі Бернуллі

$(\alpha_0, \alpha_1, \ldots, \alpha_{12})$	Точне значення	Точкова оцінка	Довірчий інтервал (побудований засобами Python)	Довірчий інтервал (побудований за формулами (2.3) і (2.4))
$(4,4,2,1,0,0,0,0,0,0,0,0,0)$	8252200	2065832	[8237749, 8257785]	[8239302, 8256233]
$(3,5,3,0,0,0,0,0,0,0,0,0,0)$	4783130	1199224	[4778883, 4794958]	[4780160, 4793683]
$(4,3,4,0,0,0,0,0,0,0,0,0,0)$	3746765	938847	[3740363, 3754794]	[3741515, 3753646]
$(3,6,1,1,0,0,0,0,0,0,0,0,0)$	3522310	883469	[3519504, 3533547]	[3520628, 3532430]
$(5,2,3,1,0,0,0,0,0,0,0,0,0)$	3274260	820441	[3268148, 3281728]	[3269236, 3280647]
$(2,7,2,0,0,0,0,0,0,0,0,0,0)$	2518615	630972	[2512622, 2524654]	[2513589, 2523695]
$(5,3,1,2,0,0,0,0,0,0,0,0,0)$	2390960	599058	[2385380, 2397120]	[2386320, 2396184]
$(4,5,1,0,1,0,0,0,0,0,0,0,0)$	2024330	507323	[2019646, 2030508]	[2020518, 2029639]
$(5,3,2,0,1,0,0,0,0,0,0,0,0)$	2017070	505279	[2011495, 2022333]	[2012364, 2021468]
$(4,5,0,2,0,0,0,0,0,0,0,0,0)$	1202135	301977	[1201164, 1209631]	[1201844, 1208956]
$(6,2,1,1,1,0,0,0,0,0,0,0,0)$	869990	217816	[865843, 873065]	[866422, 872489]
$(5,4,0,1,1,0,0,0,0,0,0,0,0)$	744150	186164	[739766, 746453]	[740302, 745920]
$(6,1,2,2,0,0,0,0,0,0,0,0,0)$	722370	180148	[715804, 722387]	[716303, 722378]
$(5,4,1,0,0,1,0,0,0,0,0,0,0)$	559020	140163	[556579, 562396]	[557047, 561933]
$(2,8,0,1,0,0,0,0,0,0,0,0,0)$	464035	116017	[460456, 465754]	[460881, 465333]
$(5,1,5,0,0,0,0,0,0,0,0,0,0)$	445280	111571	[442757, 447955]	[443177, 447543]
$(6,1,3,0,1,0,0,0,0,0,0,0,0)$	395670	99516	[394781, 399695]	[395177, 399304]
$(6,2,2,0,0,1,0,0,0,0,0,0,0)$	335775	84041	[333209, 337725]	[333571, 337367]
$(3,7,0,0,1,0,0,0,0,0,0,0,0)$	307340	76845	[304581, 308905]	[304929, 308560]
$(4,6,0,0,0,1,0,0,0,0,0,0,0)$	192390	48217	[190754, 194184]	[191030, 193912]
$(6,2,0,3,0,0,0,0,0,0,0,0,0)$	182710	45927	[181653, 185003]	[181924, 184736]
$(6,0,4,1,0,0,0,0,0,0,0,0,0)$	166375	41602	[164477, 167659]	[164731, 167409]
$(6,3,0,1,0,1,0,0,0,0,0,0,0)$	152460	38058	[150394, 153441]	[150639, 153201]
$(6,3,1,0,0,0,1,0,0,0,0,0,0)$	119790	29649	[117004, 119695]	[117222, 119485]
$(6,3,0,0,2,0,0,0,0,0,0,0,0)$	82280	20691	[81470, 83718]	[81650, 83542]
$(7,1,1,1,0,1,0,0,0,0,0,0,0)$	65340	16172	[63559, 65552]	[63721, 65394]
$(5,5,0,0,0,0,1,0,0,0,0,0,0)$	59048	14835	[58266, 60171]	[58419, 60022]
$(7,0,2,1,1,0,0,0,0,0,0,0,0)$	58080	14520	[57017, 58902]	[57170, 58757]
$(7,1,0,2,1,0,0,0,0,0,0,0,0)$	41140	10195	[39908, 41486]	[40034, 41365]
$(7,1,1,0,2,0,0,0,0,0,0,0,0)$	38720	9676	[37857, 39394]	[37980, 39276]
$(0,11,0,0,0,0,0,0,0,0,0,0,0)$	37851	9387	[36715, 38229]	[36836, 38113]
$(7,1,2,0,0,0,1,0,0,0,0,0,0)$	22990	5627	[21874, 23048]	[21971, 22999]
$(7,0,1,3,0,0,0,0,0,0,0,0,0)$	19360	4800	[18621, 19703]	[18707, 19622]
$(7,2,0,0,1,1,0,0,0,0,0,0,0)$	19360	4785	[18561, 19644]	[18648, 19561]
$(7,2,0,1,0,0,1,0,0,0,0,0,0)$	18150	4571	[17719, 18777]	[17804, 18697]
$(6,4,0,0,0,0,0,1,0,0,0,0,0)$	18150	4518	[17511, 18562]	[17595, 18482]
$(7,0,3,0,0,1,0,0,0,0,0,0,0)$	13310	3398	[13108, 14023]	[13103, 13953]
$(7,2,1,0,0,0,0,1,0,0,0,0,0)$	10890	2734	[10506, 11325]	[10572, 11263]
$(8,0,0,2,0,1,0,0,0,0,0,0,0)$	4235	1064	[3995, 4503]	[4035, 4468]
$(8,0,1,0,1,1,0,0,0,0,0,0,0)$	3630	972	[3640, 4132]	[3681, 4096]
$(8,0,0,1,2,0,0,0,0,0,0,0,0)$	3025	758	[2814, 3242]	[2847, 3213]
$(8,0,1,1,0,0,1,0,0,0,0,0,0)$	2420	635	[2339, 2735]	[2371, 2707]
$(8,0,2,0,0,0,0,1,0,0,0,0,0)$	2420	614	[2259, 2647]	[2290, 2620]
$(7,3,0,0,0,0,0,0,1,0,0,0,0)$	2420	614	[2259, 2647]	[2290, 2620]
$(8,1,0,1,0,0,0,1,0,0,0,0,0)$	1210	334	[1193, 1477]	[1205, 1460]
$(8,1,0,0,1,0,1,0,0,0,0,0,0)$	1210	326	[1161, 1445]	[1185, 1427]
$(8,1,1,0,0,0,0,0,1,0,0,0,0)$	1210	298	[1057, 1326]	[1078, 1310]
$(8,2,0,0,0,0,0,0,0,1,0,0,0)$	605	171	[582, 787]	[599, 775]
$(8,1,0,0,0,2,0,0,0,0,0,0,0)$	605	142	[475, 663]	[490, 652]
$(10,0,0,0,0,0,0,0,0,0,0,0,1)$	11	4	[3, 32]	[5, 37]

При апроксимації нормальним розподілом отримуємо наступні моменти [9]:

$$MX = Np,\ DX = Np(1-p),$$

а довірчий інтервал для імовірності отримати перестановку з характеристикою $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ з рівнем значущості α має такі границі [10]:

$$p_{\text{нижнє}} = p - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{N}},$$

$$p_{\text{верхнє}} = p + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{N}},$$

звідки легко знайти границі інтервалу для потужності з точністю до округлення до цілого числа:

$$k_{\text{нижнє}} = n! \cdot \left(p - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{N}}\right),$$

$$k_{\text{верхнє}} = n! \cdot \left(p + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{N}}\right),$$

де $z_{1-\frac{\alpha}{2}}$ — $1-\frac{\alpha}{2}$-квантиль стандартного нормального розподілу.

При апроксимації розподілом Пуассона параметр λ має наступне значення [9]:

$$\lambda = Np,$$

при чому згідно з [10] границі довірчого інтервалу для потужності класу з рівнем значущості α такі:

$$k_{\text{нижнє}} = \frac{n! \cdot \chi^2(\frac{\alpha}{2}, 2\lambda)}{2N},$$

$$k_{\text{верхнє}} = \frac{n! \cdot \chi^2(1-\frac{\alpha}{2}, 2\lambda+2)}{2N},$$

Таким чином була порахована кількість появ характеристик при генерації 10^7 випадкових перестановок при $n = 11, 26, 30, 31, 32, 33, 45, 55$ та побудовані довірчі інтервали для потужностей класів для кожного n.

У таблиці 2.4 наведені точні значення потужності усіх класів при $n = 11$ ($11! = 39916800$), точкові оцінки при генерації $N = 10^7$ випадкових перестановок і довірчі інтервали для потужності при апроксимації нормальним розподілом та розподілом Пуассона з рівнем значущості $\alpha = 0.05$.

Таблиця 1 – Довірчі інтервали для потужності усіх класів перестановок довжини $n = 11$ з використанням апроксимацій

$(\alpha_0, \alpha_1, \ldots, \alpha_{12})$	Точне значення	Точкова оцінка	Довірчий інтервал (апроксимація норм. розподілом)	Довірчий інтервал (апроксимація розподілом Пуассона)
$(4,4,2,1,0,0,0,0,0,0,0,0,0)$	8252200	2065832	[8236124, 8256157]	[8234899, 8257393]
$(3,5,3,0,0,0,0,0,0,0,0,0,0)$	4783130	1199224	[4778881, 4794956]	[4778354, 4795494]
$(4,3,4,0,0,0,0,0,0,0,0,0,0)$	3746765	938847	[3740360, 3754793]	[3740000, 3755166]
$(3,6,1,1,0,0,0,0,0,0,0,0,0)$	3522310	883469	[3519504, 3533547]	[3519175, 3533887]
$(5,2,3,1,0,0,0,0,0,0,0,0,0)$	3274260	820441	[3268148, 3281728]	[3267855, 3282033]
$(2,7,2,0,0,0,0,0,0,0,0,0,0)$	2518615	630972	[2512623, 2524654]	[2512427, 2524861]
$(5,3,1,2,0,0,0,0,0,0,0,0,0)$	2390960	599058	[2385376, 2397119]	[2385196, 2397311]
$(4,5,1,0,1,0,0,0,0,0,0,0,0)$	2024330	507323	[2019645, 2030505]	[2019506, 2030656]
$(5,3,2,0,1,0,0,0,0,0,0,0,0)$	2017070	505279	[2011493, 2022331]	[2011354, 2022482]
$(4,5,0,2,0,0,0,0,0,0,0,0,0)$	1202135	301977	[1201161, 1209630]	[1201100, 1209703]
$(6,2,1,1,1,0,0,0,0,0,0,0,0)$	869990	217816	[865840, 873064]	[865804, 873111]
$(5,4,0,1,1,0,0,0,0,0,0,0,0)$	744150	186164	[739763, 746452]	[739735, 746491]
$(6,1,2,2,0,0,0,0,0,0,0,0,0)$	722370	180148	[715802, 722384]	[715776, 722422]
$(5,4,1,0,0,1,0,0,0,0,0,0,0)$	559020	140163	[556577, 562395]	[556560, 562423]
$(2,8,0,1,0,0,0,0,0,0,0,0,0)$	464035	116017	[460453, 465753]	[460441, 465776]
$(5,1,5,0,0,0,0,0,0,0,0,0,0)$	445280	111571	[442757, 447955]	[442746, 447977]
$(6,1,3,0,1,0,0,0,0,0,0,0,0)$	395670	99516	[394780, 399692]	[394771, 399712]
$(6,2,2,0,0,1,0,0,0,0,0,0,0)$	335775	84041	[333206, 337724]	[333200, 337741]
$(3,7,0,0,1,0,0,0,0,0,0,0,0)$	307340	76845	[304580, 308902]	[304575, 308918]
$(4,6,0,0,0,1,0,0,0,0,0,0,0)$	192390	48217	[190753, 194181]	[190752, 194193]
$(6,2,0,3,0,0,0,0,0,0,0,0,0)$	182710	45927	[181653, 184999]	[181653, 185011]
$(6,0,4,1,0,0,0,0,0,0,0,0,0)$	166375	41602	[164473, 167659]	[164473, 167670]
$(6,3,0,1,0,1,0,0,0,0,0,0,0)$	152460	38058	[150392, 153439]	[150392, 153450]
$(6,3,1,0,0,0,1,0,0,0,0,0,0)$	119790	29649	[117004, 119695]	[117005, 119705]
$(6,3,0,0,2,0,0,0,0,0,0,0,0)$	82280	20691	[81467, 83717]	[81470, 83726]
$(7,1,1,1,0,1,0,0,0,0,0,0,0)$	65340	16172	[63559, 65548]	[63562, 65557]
$(5,5,0,0,0,0,1,0,0,0,0,0,0)$	59048	14835	[58264, 60169]	[58267, 60178]
$(7,0,2,1,1,0,0,0,0,0,0,0,0)$	58080	14520	[57017, 58902]	[57020, 58910]
$(7,1,0,2,1,0,0,0,0,0,0,0,0)$	41140	10195	[39905, 41485]	[39909, 41493]
$(7,1,1,0,2,0,0,0,0,0,0,0,0)$	38720	9676	[37854, 39393]	[37857, 39401]
$(0,11,0,0,0,0,0,0,0,0,0,0,0)$	37851	9387	[36712, 38228]	[36715, 38236]
$(7,1,2,0,0,0,1,0,0,0,0,0,0)$	22990	5627	[21874, 23048]	[21878, 23056]
$(7,0,1,3,0,0,0,0,0,0,0,0,0)$	19360	4800	[18618, 19702]	[18621, 19710]
$(7,2,0,0,1,1,0,0,0,0,0,0,0)$	19360	4785	[18559, 19642]	[18562, 19650]
$(7,2,0,1,0,0,1,0,0,0,0,0,0)$	18150	4571	[17717, 18775]	[17720, 18783]
$(6,4,0,0,0,0,0,1,0,0,0,0,0)$	18150	4518	[17508, 18561]	[17512, 18569]
$(7,0,3,0,0,1,0,0,0,0,0,0,0)$	13310	3398	[13107, 14020]	[13111, 14028]
$(7,2,1,0,0,0,0,1,0,0,0,0,0)$	10890	2734	[10504, 11323]	[10507, 11331]
$(8,0,0,2,0,1,0,0,0,0,0,0,0)$	4235	1064	[3991, 4503]	[3995, 4511]
$(8,0,1,0,1,1,0,0,0,0,0,0,0)$	3630	972	[3639, 4128]	[3643, 4136]
$(8,0,0,1,2,0,0,0,0,0,0,0,0)$	3025	758	[2810, 3242]	[2814, 3249]
$(8,0,1,1,0,0,1,0,0,0,0,0,0)$	2420	635	[2337, 2732]	[2341, 2740]
$(8,0,2,0,0,0,0,1,0,0,0,0,0)$	2420	614	[2257, 2645]	[2260, 2653]
$(7,3,0,0,0,0,0,0,1,0,0,0,0)$	2420	614	[2257, 2645]	[2260, 2653]
$(8,1,0,1,0,0,0,1,0,0,0,0,0)$	1210	334	[1190, 1477]	[1194, 1485]
$(8,1,0,0,1,0,1,0,0,0,0,0,0)$	1210	326	[1160, 1443]	[1163, 1451]
$(8,1,1,0,0,0,0,0,1,0,0,0,0)$	1210	298	[1054, 1325]	[1058, 1333]
$(8,2,0,0,0,0,0,0,0,1,0,0,0)$	605	171	[580, 785]	[584, 793]
$(8,1,0,0,0,2,0,0,0,0,0,0,0)$	605	142	[473, 661]	[477, 669]
$(10,0,0,0,0,0,0,0,0,0,0,0,1)$	11	4	[0, 32]	[4, 41]

При $n = 11$ усі можливі перестановки діляться на 50 класів за ознакою вигляду характеристики. Як можна бачити у таблиці 2.4, перестановок без перепайок (з характеристикою $(0, 11, 0, 0, 0, 0, 0, 0, 0, 0, 0)$) існує 37851, що становить 0.000948247354497 від усіх можливих перестановок. Щодо перестановок з виродженою характеристикою, їх усього 11 (0.000000275573192 від усіх перестановок).

Отримані довірчі інтервали при апроксимації біноміального розподілу нормальних розподілом та розподілом Пуассона виявилися досить близькими, при чому чим менша потужність класу, тим більше вони збігаються. Для класів з більшою потужністю інтервали при апроксимації розподілом Пуассона ширші і перекривають інтервали, отримані при апроксимації нормальним розподілом. Це дає підстави вважати, що обидва інтервали були побудовані коректно. Довірчі інтервали, побудовані засобами Python, майже співпадають з інтервалами при апроксимації нормальним розподілом.

Лише два точних значення потужності класів з 50-ти не потрапили до довірчих інтервалів, що допускається 95-відсотковим інтервалом, оскільки $\frac{2}{50} = 0.04 < 0.05$. Так, кількість перестановок із характеристикою $(6, 3, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0)$ дорівнює 119790, що на 95 більше крайнього значення довірчого інтервалу при використанні засобів Python, на 305 для біноміальної моделі, на 95 при апроксимації нормальним розподілом та на 85 – при апроксимації розподілом Пуассона. Другий клас, чия характеристика не потрапила до інтервалу — це $(8, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0)$. Його дійсна потужніть становить 3630, а найменші значення інтервалів становлять 3640 за використання засобів Python, 3681 для біноміальної моделі, 3639 при апроксимації нормальним розподілом та 3643 при апроксимації розподілом Пуассона. Для усіх інтервалів похибка становить менше 100 перестановок, що дуже мало в порівнянні з кількістю усіх перестановок довжини $n = 11$. Оскільки точні значення обох характеристик не потрапили до жодного з побудованих інтервалів, можна зробити припущення, що в даній вибірці 10^7

випадкових перестановок є деяка аномалія, за рахунок якої перестановок з характеристикою $(6, 3, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0)$ отримано менше, ніж мало б, а з характеристикою $(8, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0)$ — більше. У наступній вибірці така аномалія для цих класів не проявиться, проте можлива для інших класів. Подібних похибок можна було б уникнути більшим об'ємом вибірки випадкових перестановок за наявності необхідних обчислюваних потужностей.

Загалом, можна сказати, що усі використані способи побудови довірчих інтервалів виявилися досить дієвими у випадку оцінки потужностей класів перестановок довжини $n = 11$, проте найбільш ефективним є інтервал, отриманий за формулами (2.3) і (2.4) для біноміальної моделі, оскільки він є найбільш вузьким і дає точніше уявлення про дійсне значення потужності класу. Спосіб побудови інтервалів для подальших досліджень можна обирати в залежності від наявних обчислювальних ресурсів і необхідної точності інтервалів.

Застосуємо описані алгоритми для побудови довірчих інтервалів для перестановок більшої довжини.

Зважаючи на те, що в англійському алфавіті 26 літер, а в українському та російському — 33, наведемо деякі класи та довірчі інтервали для їх потужності для перестановок довжин $n = 26$ ($26! \approx 4.0329 \cdot 10^{26}$) та $n = 33$ ($33! \approx 8.6833 \cdot 10^{36}$) (таблиці 2.5 та 2.6 відповідно).

Також отримані довірчі інтервали для потужності класів перестановок для довжин $n = 30, 31, 32, 45, 55$. У таблицях 2.7, 2.8, 2.9, 2.10 та 2.11 відповідно наведені деякі класи разом з точковими оцінками та довірчими інтервалами, побудованими за допомогою формул (2.3) і (2.4), апроксимації нормальним розподілом та розподілом Пуассона.

Таблиця 2.5 – Довірчі інтервали для класів перестановок довжини $n = 26$

$(\alpha_0, \alpha_1, \ldots, \alpha_{27})$	Точкова оцінка	Довірчий інтервал (побудований за формулами (2.3) і (2.4))	Довірчий інтервал (апроксимація норм. розподілом)	Довірчий інтервал (апроксимація розподілом Пуассона)
$(9, 10, 5, 2, 0, \ldots, 0)$	676522	$[2.723086 \cdot 10^{25}, 2.73363 \cdot 10^{25}]$	$[2.722078 \cdot 10^{25}, 2.73463 \cdot 10^{25}]$	$[2.72186 \cdot 10^{25}, 2.73486 \cdot 10^{25}]$
$(10, 8, 6, 2, 0, \ldots, 0)$	526935	$[2.120397 \cdot 10^{25}, 2.12978 \cdot 10^{25}]$	$[2.11950 \cdot 10^{25}, 2.13067 \cdot 10^{25}]$	$[2.11935 \cdot 10^{25}, 2.13083 \cdot 10^{25}]$
$(10, 9, 4, 3, 0, \ldots, 0)$	406568	$[1.63551 \cdot 10^{25}, 1.64380 \cdot 10^{25}]$	$[1.63472 \cdot 10^{25}, 1.64459 \cdot 10^{25}]$	$[1.63462 \cdot 10^{25}, 1.64470 \cdot 10^{25}]$
$(7, 13, 5, 1, 0, \ldots, 0)$	304227	$[1.22332 \cdot 10^{25}, 1.23053 \cdot 10^{25}]$	$[1.22263 \cdot 10^{25}, 1.23121 \cdot 10^{25}]$	$[1.22257 \cdot 10^{25}, 1.23129 \cdot 10^{25}]$
$(11, 7, 6, 1, 1, 0, \ldots, 0)$	215113	$[8.64490 \cdot 10^{24}, 8.70582 \cdot 10^{24}]$	$[8.63906 \cdot 10^{24}, 8.71159 \cdot 10^{24}]$	$[8.63870 \cdot 10^{24}, 8.71206 \cdot 10^{24}]$
$(7, 14, 3, 2, 0, 0, \ldots, 0)$	150953	$[6.06225 \cdot 10^{24}, 6.11344 \cdot 10^{24}]$	$[6.05733 \cdot 10^{24}, 6.11828 \cdot 10^{24}]$	$[6.05713 \cdot 10^{24}, 6.11859 \cdot 10^{24}]$
$(9, 12, 2, 2, 1, 0, \ldots, 0)$	115477	$[4.63470 \cdot 10^{24}, 4.67956 \cdot 10^{24}]$	$[4.63038 \cdot 10^{24}, 4.68380 \cdot 10^{24}]$	$[4.63026 \cdot 10^{24}, 4.68403 \cdot 10^{24}]$
$(6, 14, 6, 0, 0, 0, \ldots, 0)$	70359	$[2.820007 \cdot 10^{24}, 2.85511 \cdot 10^{24}]$	$[2.81663 \cdot 10^{24}, 2.85841 \cdot 10^{24}]$	$[2.81659 \cdot 10^{24}, 2.85856 \cdot 10^{24}]$
$(11, 5, 9, 1, 0, 0, \ldots, 0)$	47858	$[1.91562 \cdot 10^{24}, 1.94461 \cdot 10^{24}]$	$[1.91282 \cdot 10^{24}, 1.94732 \cdot 10^{24}]$	$[1.91282 \cdot 10^{24}, 1.94744 \cdot 10^{24}]$
$(11, 9, 1, 5, 0, 0, \ldots, 0)$	11032	$[4.37970 \cdot 10^{23}, 4.51938 \cdot 10^{23}]$	$[4.36614 \cdot 10^{23}, 4.53209 \cdot 10^{23}]$	$[4.36647 \cdot 10^{23}, 4.53292 \cdot 10^{23}]$
$(6, 17, 1, 1, 1, 0, \ldots, 0)$	3268	$[1.28067 \cdot 10^{23}, 1.35692 \cdot 10^{23}]$	$[1.27317 \cdot 10^{23}, 1.36355 \cdot 10^{23}]$	$[1.27355 \cdot 10^{23}, 1.36434 \cdot 10^{23}]$
$(3, 20, 3, 0, 0, 0, \ldots, 0)$	226	$[1.40603 \cdot 10^{21}, 1.01765 \cdot 10^{22}]$	$[7.92611 \cdot 10^{21}, 1.03027 \cdot 10^{22}]$	$[7.96474 \cdot 10^{21}, 1.0383 \cdot 10^{22}]$
$(15, 4, 1, 5, 0, 1, 0, \ldots, 0)$	65	$[2.11047 \cdot 10^{21}, 3.22257 \cdot 10^{21}]$	$[1.98413 \cdot 10^{21}, 3.25866 \cdot 10^{21}]$	$[2.02314 \cdot 10^{21}, 3.34118 \cdot 10^{21}]$
$(16, 2, 1, 6, 1, 0, \ldots, 0)$	7	$[1.32494 \cdot 10^{20}, 5.30252 \cdot 10^{20}]$	$[7.31742 \cdot 10^{20}, 4.91434 \cdot 10^{20}]$	$[1.13501 \cdot 10^{20}, 5.81654 \cdot 10^{20}]$
$(17, 1, 5, 1, 0, 1, 0, 1, 0, \ldots)$	1	$[2.06861 \cdot 10^{18}, 1.91316 \cdot 10^{20}]$	$[0, 1.19373 \cdot 10^{20}]$	$[1.02105 \cdot 10^{18}, 2.24699 \cdot 10^{20}]$

Таблиця 2.6 – Довірчі інтервали для класів перестановок довжини $n = 33$

$(\alpha_0, \alpha_1, \ldots, \alpha_{34})$	Точкова оцінка	Довірчий інтервал (побудований за формулами (2.3) і (2.4))	Довірчий інтервал (апроксимація норм. розподілом)	Довірчий інтервал (апроксимація розподілом Пуассона)
$(11, 13, 7, 2, 0, \ldots, 0)$	458246	$[3.96965 \cdot 10^{35}, 3.98856 \cdot 10^{35}]$	$[3.96784 \cdot 10^{35}, 3.99035 \cdot 10^{35}]$	$[3.96758 \cdot 10^{35}, 3.99063 \cdot 10^{35}]$
$(12, 12, 7, 1, 1, 0, \ldots, 0)$	272792	$[2.36139 \cdot 10^{35}, 2.37611 \cdot 10^{35}]$	$[2.35998 \cdot 10^{35}, 2.37751 \cdot 10^{35}]$	$[2.359859 \cdot 10^{35}, 2.37765 \cdot 10^{35}]$
$(12, 10, 10, 1, 0, \ldots, 0)$	136820	$[1.18281 \cdot 10^{35}, 1.19331 \cdot 10^{35}]$	$[1.18180 \cdot 10^{35}, 1.19430 \cdot 10^{35}]$	$[1.18176 \cdot 10^{35}, 1.19436 \cdot 10^{35}]$
$(10, 17, 3, 2, 1, 0, \ldots, 0)$	52903	$[4.56102 \cdot 10^{34}, 4.62664 \cdot 10^{35}]$	$[4.55469 \cdot 10^{34}, 4.63278 \cdot 10^{35}]$	$[4.55467 \cdot 10^{34}, 4.63305 \cdot 10^{35}]$
$(9, 18, 4, 1, 1, 0, \ldots, 0)$	41219	$[3.55029 \cdot 10^{34}, 3.60825 \cdot 10^{34}]$	$[3.54466 \cdot 10^{34}, 3.61366 \cdot 10^{34}]$	$[3.54471 \cdot 10^{34}, 3.61390 \cdot 10^{34}]$
$(15, 8, 7, 2, 0, 1, 0, \ldots, 0)$	24644	$[2.11757 \cdot 10^{34}, 2.16245 \cdot 10^{34}]$	$[2.11323 \cdot 10^{34}, 2.16660 \cdot 10^{34}]$	$[2.11328 \cdot 10^{34}, 2.16680 \cdot 10^{34}]$
$(9, 19, 2, 2, 1, 0, \ldots, 0)$	8756	$[7.470017 \cdot 10^{33}, 7.73807 \cdot 10^{33}]$	$[7.44393 \cdot 10^{33}, 7.76230 \cdot 10^{33}]$	$[7.44469 \cdot 10^{33}, 7.76406 \cdot 10^{33}]$
$(13, 7, 13, 0, \ldots, 0)$	3388	$[2.85928 \cdot 10^{33}, 3.02640 \cdot 10^{33}]$	$[2.84286 \cdot 10^{33}, 3.04095 \cdot 10^{33}]$	$[28.43671 \cdot 10^{33}, 3.04267 \cdot 10^{33}]$
$(11, 17, 2, 2, 0, 0, 1, 0, \ldots, 0)$	1187	$[9.82003 \cdot 10^{32}, 1.08129 \cdot 10^{33}]$	$[9.72078 \cdot 10^{33}, 1.08934 \cdot 10^{33}]$	$[9.72901 \cdot 10^{32}, 1.09106 \cdot 10^{33}]$
$(18, 6, 4, 3, 0, 2, 0, \ldots, 0)$	142	$[1.06791 \cdot 10^{32}, 1.41730 \cdot 10^{32}]$	$[1.03023 \cdot 10^{32}, 1.43583 \cdot 10^{32}]$	$[1.03857 \cdot 10^{32}, 1.45333 \cdot 10^{32}]$
$(19, 5, 3, 4, 0, 2, 0, \ldots, 0)$	16	$[8.71454 \cdot 10^{30}, 2.11015 \cdot 10^{31}]$	$[7.08572 \cdot 10^{30}, 2.07009 \cdot 10^{31}]$	$[7.94123 \cdot 10^{30}, 2.25619 \cdot 10^{31}]$
$(9, 21, 0, 2, 0, 0, 1, 0, \ldots, 0)$	5	$[1.71074 \cdot 10^{30}, 9.12880 \cdot 10^{30}]$	$[5.36098 \cdot 10^{30}, 8.14722 \cdot 10^{30}]$	$[1.40972 \cdot 10^{30}, 1.01320 \cdot 10^{31}]$
$(20, 3, 2, 6, 2, 0, \ldots, 0)$	1	$[4.45396 \cdot 10^{28}, 4.11925 \cdot 10^{30}]$	$[0, 2.57023 \cdot 10^{30}]$	$[2.19843 \cdot 10^{28}, 4.83803 \cdot 10^{30}]$

Таблиця 2.7 – Довірчі інтервали для класів перестановок довжини $n = 30$

$(\alpha_0, \alpha_1, \ldots, \alpha_{31})$	Точкова оцінка	Довірчий інтервал (побудований за формулами (2.3) і (2.4))	Довірчий інтервал (апроксимація норм. розподілом)	Довірчий інтервал (апроксимація розподілом Пуассона)
$(10, 12, 6, 2, 0, \ldots, 0)$	539643	$[1.4283 \cdot 10^{31}, 1.4345 \cdot 10^{31}]$	$[1.4277 \cdot 10^{31}, 1.4351 \cdot 10^{31}]$	$[1.4276 \cdot 10^{31}, 1.4352 \cdot 10^{31}]$
$(7, 18, 3, 2, 0, \ldots, 0)$	25261	$[6.6314 \cdot 10^{29}, 6.7702 \cdot 10^{29}]$	$[6.6180 \cdot 10^{29}, 6.7831 \cdot 10^{29}]$	$[6.6181 \cdot 10^{29}, 6.7837 \cdot 10^{29}]$
$(16, 6, 3, 2, 3, 0, \ldots, 0)$	1181	$[2.9842 \cdot 10^{28}, 3.2868 \cdot 10^{28}]$	$[2.9540 \cdot 10^{28}, 3.3113 \cdot 10^{28}]$	$[2.9565 \cdot 10^{28}, 3.3165 \cdot 10^{28}]$
$(5, 23, 0, 1, 1, 0, \ldots, 0)$	8	$[1.0559 \cdot 10^{26}, 3.8288 \cdot 10^{26}]$	$[6.5156 \cdot 10^{25}, 3.5925 \cdot 10^{26}]$	$[9.1614 \cdot 10^{25}, 4.1812 \cdot 10^{26}]$

Таблиця 2.8 – Довірчі інтервали для класів перестановок довжини $n = 31$

$(\alpha_0, \alpha_1, \ldots, \alpha_{32})$	Точкова оцінка	Довірчий інтервал (побудований за формулами (2.3) і (2.4))	Довірчий інтервал (апроксимація норм. розподілом)	Довірчий інтервал (апроксимація розподілом Пуассона)
$(11, 11, 7, 2, 0, \ldots, 0)$	495301	$[4.0635 \cdot 10^{32}, 4.0821 \cdot 10^{32}]$	$[4.0617 \cdot 10^{32}, 4.0838 \cdot 10^{32}]$	$[4.0614 \cdot 10^{32}, 4.0841 \cdot 10^{32}]$
$(7, 17, 7, 0, \ldots, 0)$	26362	$[2.1458 \cdot 10^{31}, 2.1898 \cdot 10^{31}]$	$[2.1416 \cdot 10^{31}, 2.1938 \cdot 10^{31}]$	$[2.1416 \cdot 10^{31}, 2.1940 \cdot 10^{31}]$
$(8, 19, 2, 0, 2, 0, \ldots, 0)$	866	$[6.7277 \cdot 10^{29}, 7.5321 \cdot 10^{29}]$	$[6.6467 \cdot 10^{29}, 7.5952 \cdot 10^{29}]$	$[6.6545 \cdot 10^{29}, 7.6115 \cdot 10^{29}]$
$(8, 20, 0, 2, 0, 1, 0, \ldots, 0)$	63	$[4.1556 \cdot 10^{28}, 6.3893 \cdot 10^{28}]$	$[3.9012 \cdot 10^{28}, 6.4596 \cdot 10^{28}]$	$[3.9806 \cdot 10^{28}, 6.6280 \cdot 10^{28}]$

Таблиця 2.9 – Довірчі інтервали для класів перестановок довжини $n = 32$

$(\alpha_0, \alpha_1, \ldots, \alpha_{33})$	Точкова оцінка	Довірчий інтервал (побудований за формулами (2.3) і (2.4))	Довірчий інтервал (апроксимація норм. розподілом)	Довірчий інтервал (апроксимація розподілом Пуассона)
$(11, 12, 7, 2, 0, \ldots, 0)$	489821	$[1.2859 \cdot 10^{34}, 1.2918 \cdot 10^{34}]$	$[1.2854 \cdot 10^{34}, 1.2924 \cdot 10^{34}]$	$[1.2853 \cdot 10^{34}, 1.2925 \cdot 10^{34}]$
$(10, 12, 10, 0, \ldots, 0)$	61211	$[1.6000 \cdot 10^{33}, 1.6214 \cdot 10^{33}]$	$[1.5979 \cdot 10^{33}, 1.6234 \cdot 10^{33}]$	$[1.5979 \cdot 10^{33}, 1.6235 \cdot 10^{33}]$
$(7, 20, 3, 2, 0, \ldots, 0)$	9090	$[2.3508 \cdot 10^{32}, 2.4335 \cdot 10^{32}]$	$[2.3427 \cdot 10^{32}, 2.4410 \cdot 10^{32}]$	$[2.3429 \cdot 10^{32}, 2.4415 \cdot 10^{32}]$
$(19, 3, 4, 3, 3, 0, \ldots, 0)$	34	$[6.5809 \cdot 10^{29}, 1.1911 \cdot 10^{30}]$	$[5.9393 \cdot 10^{29}, 1.1954 \cdot 10^{30}]$	$[6.1957, 1.25018 \cdot 10^{30}]$

Таблиця 2.10 – Довірчі інтервали для класів перестановок довжини $n = 45$

$(\alpha_0, \alpha_1, \ldots, \alpha_{46})$	Точкова оцінка	Довірчий інтервал (побудований за формулами (2.3) і (2.4))	Довірчий інтервал (апроксимація норм. розподілом)	Довірчий інтервал (апроксимація розподілом Пуассона)
$(15, 18, 9, 3, 0, \ldots, 0)$	259946	$[3.0996 \cdot 10^{54}, 3.1195 \cdot 10^{54}]$	$[3.0977 \cdot 10^{54}, 3.1213 \cdot 10^{54}]$	$[3.0976 \cdot 10^{54}, 3.1215 \cdot 10^{54}]$
$(21, 11, 8, 2, 3, 0, \ldots, 0)$	5668	$[6.6328 \cdot 10^{52}, 6.9302 \cdot 10^{52}]$	$[6.6037 \cdot 10^{52}, 6.9566 \cdot 10^{52}]$	$[6.6048 \cdot 10^{52}, 6.9590 \cdot 10^{52}]$
$(11, 26, 7, 0, 0, 1, 0, \ldots, 0)$	402	$[4.4212 \cdot 10^{51}, 5.2224 \cdot 10^{51}]$	$[4.3387 \cdot 10^{51}, 5.278 \cdot 10^{51}]$	$[4.3502 \cdot 10^{51}, 5.3027]$
$(15, 23, 5, 0, 0, 1, 0, 1, 0, \ldots, 0)$	10	$[6.4900 \cdot 10^{50}, 2.0291 \cdot 10^{51}]$	$[4.5481 \cdot 10^{50}, 1.9376 \cdot 10^{51}]$	$[5.7364 \cdot 10^{50}, 2.2000 \cdot 10^{51}]$

Таблиця 2.11 – Довірчі інтервали для класів перестановок довжини $n = 55$

$(\alpha_0, \alpha_1, \ldots, \alpha_{56})$	Точкова оцінка	Довірчий інтервал (побудований за формулами (2.3) і (2.4))	Довірчий інтервал (апроксимація норм. розподілом)	Довірчий інтервал (апроксимація розподілом Пуассона)
$(20, 20, 11, 3, 1, 0, \ldots, 0)$	198890	$[2.5160 \cdot 10^{71}, 2.5344 \cdot 10^{71}]$	$[2.5142 \cdot 10^{71}, 2.5362 \cdot 10^{71}]$	$[2.5141 \cdot 10^{71}, 2.5363 \cdot 10^{71}]$
$(17, 28, 6, 2, 1, 1, 0, \ldots, 0)$	3030	$[3.7328 \cdot 10^{69}, 3.9640 \cdot 10^{69}]$	$[3.7101 \cdot 10^{69}, 3.9840 \cdot 10^{69}]$	$[3.7112 \cdot 10^{69}, 3.9865 \cdot 10^{69}]$
$(16, 29, 7, 0, 3, 0, \ldots, 0)$	285	$[3.2733 \cdot 10^{68}, 3.9914 \cdot 10^{68}]$	$[3.1984 \cdot 10^{68}, 4.0386 \cdot 10^{68}]$	$[3.2105 \cdot 10^{68}, 4.0639 \cdot 10^{68}]$
$(30, 9, 7, 5, 3, 1, 0, \ldots, 0)$	16	$[1.2742 \cdot 10^{67}, 3.0854 \cdot 10^{67}]$	$[1.0360 \cdot 10^{67}, 3.0268 \cdot 10^{67}]$	$[1.1611 \cdot 10^{67}, 3.9891 \cdot 10^{67}]$

2.7 Аналіз результатів

Як вже зазначалося раніше, кількість так званих «перепайок» у роторі, а як наслідок і в перестановці, впливає на якість шифрування роторними шифраторами і на стійкість шифру. Їх кількість можна легко визначити за виглядом характеристики $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ перестановки довжини n. Найкращими для шифрування вважаються перестановки «без перепайок», оскільки за їх використання криптоаналіз шифру є найбільш ускладненим і вимагає більше часу і ресурсів. Проте із отриманих в процесі роботи даних можна бачити, що кількість таких перестановок є малою порівняно з загальною кількістю перестановок. Так, при $n = 11$ частка перестановок «без перепайок» становить 0.000948247354497, при $n = 13$ — 0.000165467, при $n = 19$ — 0.000000720595, а при більших n ця частка майже рівна 0. Через це доводиться також використовувати перестановки, які є не такими ефективними з точки зору криптографії, але тим не менш задовольняють обраний рівень стійкості шифру. Тому важливо знати потужність класів перестановок, щоб приблизно оцінювати імовірність при випадковому виборі перестановки обрати задовільну.

Для роторних шифраторів важлива кількість унікальних значень результатів j_k суми за модулем

$$\begin{array}{cccccc} & 0 & 1 & \ldots & n-1 \\ \boxplus & i_0 & i_1 & \ldots & i_{n-1} \\ \hline & j_0 & j_1 & \cdots & j_{n-1} \end{array}$$

тому чим більше значення α_1 характеристики перестановки, тим стійкішим буде результат шифрування. В залежності від набору $(j_0, j_1, \ldots, j_{n-1})$ криптоаналіз може виявитися більш або менш складним, що добре описується у [13] для випадків з різною кількістю роторів.

Всього класів для перестановок з ненульовою точковою оцінкою довжини $n = 26$ було отримано 688, для $n = 30$ — 976, для $n = 31$ — 1065, для $n = 32$ — 1153, для $n = 33$ — 1229, для $n = 45$ — 2508, для $n = 55$ — 3906. При жодній з цих довжин перестановки без перепайок та перестановки з виродженою характеристикою при 10^7 генераціях випадкових перестановок не зустрілися. При досить великих довжинах перестановок n та при малих потужностях класів спостерігається велика розбіжність у інтервалах при апроксимації нормальним розподілом та розподілом Пуассона, проте імовірність того, що значення потужності класу лежить у межах від крайнього лівого значення інтервалу при апроксимації нормальним розподілом до крайнього правого значення при апроксимації розподілом Пуассона, прямує до 1. Враховуючи, що такий інтервал є нехтовно малим у порівнянні з загальною кількістю перестановок, то можна сказати, що при такому переході точність оцінки потужності класу зменшується не значно.

При великих n інтервали, отримані за допомогою апроксимації біноміального розподілу нормальним розподілом, переважно виявилися більш вузькими, ніж інтервали, отримані за допомогою апроксимації розподілом Пуассона, а отже дають більш точні оцінки потужностей класів принаймні при виконанні умов задовільності переходу від біноміального розподілу до нормального розподілу:

1) $Np(1-p) > 5$ і $0.1 \leq p \leq 0.9$

або

2) $Np(1-p) > 25$ для довільного p.

У наведених таблицях 2.5, 2.7, 2.8, 2.9, 2.6, 2.10, 2.11 можна бачити, що для усіх n серед α_i отриманих характеристик $(\alpha_0, \alpha_1, \ldots, \alpha_{n-1})$ є значна кількість нулів, при чому вони зосереджені у лівій частині характеристик. Це означає, що великої кількості «перепайок» немає.

Висновки до розділу 2

У розділі розглянуті загальні методи статистичного оцінювання параметрів розподілів, зокрема формули для отримання довірчих інтервалів для середнього та дисперсії біноміального розподілу, нормального розподілу та розподілу Пуассона. Проведена класифікація класів перестановок зі спеціальними властивостями, методом Монте-Карло з використанням апроксимації біноміального розподілу нормальним розподілом та розподілом Пуассона отримано інтервальні оцінки для потужності класів для довжин перестановки $n = 11, 26, 30, 31, 32, 33, 45, 55$. Наведено повний список класів перестановок з точним значенням і довірчими інтервалами для їх потужності для перестановок довжини $n = 11$. Порівняні два способи побудови довірчих інтервалів для потужностей класів. Характеристики та властивості кожного класу впливають на стійкість роторних шифрів та шифрів, які використовують як базові елементи перестановки на деяких алфавітах. Властивості перестановок та їх кількість важливі для побудові таких шифрів та проведення криптоаналізу. Як продовження досліджень планується отримати оцінки інтервалів для всіх класів спеціальних перестановок для інших значень n до 60 та показати значення властивостей класів для криптоаналізу.

ВИСНОВКИ

В роботі приведені теоретичні відомості про шифрувальну машину Енігму, теоретичні відомості з теорії імовірності, математичної статистики та методу Монте-Карло. Введено поняття характеристики перестановок, які використовуються як підключі у роторних шифрувальних машинах. У результаті роботи було виділено деякі класи перестановок за допомогою розроблених алгоритмів 1, 2, 3 та 4 і побудовані довірчі інтервали для отриманих класів. Оглянуто методи обробки даних, отриманих внаслідок генерації перестановок статистичним моделюванням. Оскільки характеристика перестановки впливає на ефективність та стійкість при їх застосування у роторних шифрувальних машинах, то актуальною задачею є оцінка потужності різних класів перестановок. Проведено статистичне моделювання та наведено результати експериментів для знаходження оцінок потужності виділених класів перестановок різного порядку. Показано, що імовірність випадково вибрати перестановку з високими криптографічними властивостями швидко зменшується з ростом порядку перестановки.

Для використання в криптографічних системах необхідно розробляти ефективні алгоритми для генерування перестановок з різними криптографічними характеристиками. У подальших дослідженнях необхідно виявити та описати інші класи і побудувати довірчі інтервали для оцінок їх потужностей за допомогою апроксимації біноміального розподілу до нормального розподілу та розподілу Пуассона.

ПЕРЕЛІК ПОСИЛАНЬ

1. Ивченко Г. И. Математическая статистика / Г. И. Ивченко, Ю. И. Медведев. – Москва: Высшая школа, 1984. – 248 с.

2. Кнут Д. Искусство программирования / Д. Кнут. – Москва: Вильямс, 2010. – 720 с. – (3).

3. Ивченко Г. И. Введение в математическую статистику / Г. И. Ивченко, Ю. И. Медведев. – Москва: ЛКИ, 2010. – 600 с.

4. Коваленко И.Н. Об одной верхней оценке числа полных отображений / Кибернетика и системный анализ. – 1996. – № 1. – С. 81-85.

5. Коваленко І.М., Купер К. Верхня границя для числа повних відображень / Теорія ймовірностей і математична статистика. – 1995. – Т. 53. – С. 69-75

6. Cooper C., Gilchrist R., Kovalenko I.N., Novacovic D. Deriving the number of good permutations with applications to cryptography / Кибернетика и системный анализ. – 1999. – № 5. – P. 10-16

7. Кузнецов Н.Ю. Применение ускоренного моделирования к нахождению количества “хороших” перестановок / Кибернетика и системный анализ. – 2007. – № 6. – С. 80-89.

8. Ширяєв А. Н. Вероятность / А. Н. Ширяєв. – Москва: МЦНМО, 2007. – 552 с. – (4).

9. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. / А. И. Кобзарь. – Москва: ФИЗМАТЛИТ, 2006. – 816 с.

10. Шор Я. Б. Таблицы для анализа и контроля надежности / Я. Б. Шор, Ф. И. Кузьмин. – Москва: Советское радио, 1968. – 288 с.

11. Я.Б.Шор. Статистические методы анализа и контроля качества и надежности. – Москва: Сов. Радио, 1962.- 552 с.

12. Матеріали XVII Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених «Теоретичні і прикладні проблеми

фізики, математики та інформатики» (25-26 квітня 2019 р., м. Київ, Україна)/ Уклад.: Василенко О. Д., Пономаренко С. М., Бех С. В., Степаненко В. М., Мирошникова І. Ю., Козленко О. В., Кірієнко О. В., Яковлєв С. В., Деркач О. Г. - Київ: ВПІ ВПК «ПОЛІТЕХНІКА», 2019. - С. 195-196.

13. Konheim A. G. Computer security and cryptography / Alan G. Konheim. – New Jersey: John Wiley and Sons, Inc., 2007. – 521 с.

ДОДАТОК А ТЕКСТИ ПРОГРАМ

Тексти інструментальних програм, використовуваних у дослідженні.

А.1 Код програми для визначення кількості перестановок «без перепайок»

```
import itertools
import csv
from collections import Counter

def res(perms):
  modsumm=[]
  for perm in perms:
    i=0
    p=[]
    while i<n:
      p.append((perm[i]+arr[i])%n)
      i+=1
    modsumm.append(p)
  return modsumm

print('n'.center(5), 'Without solderings'.center(15))
table=[]

for n in range(13, 14):
  if n%2 != 0:
    arr = range(n)
```

```
    perms = list(itertools.permutations(arr))
    modsumms = res(perms)

    withoutsoldering = []

    for modsumm in modsumms:
      if len(set(modsumm)) == len(modsumm):
        withoutsoldering.append(modsumm)
    l=[n, len(withoutsoldering)]
    table.append(l)

data = open('count_of_solderings_13.txt', 'w')
data.write('n'.center(5))
data.write('Without solderings'.center(15)+'\n')

for i in table:
  print(str(i[0]).center(5), str(i[1]).center(20))
  data.write(str(i[0]).center(5))
  data.write(str(i[1]).center(20)+'\n')

data.close()
```

.

A.2 Код програми для повного перебору перестановок довжини 11 та виділення класів

```
import itertools

def res(perm):
  modsumm=[]
  i=0
  while i<n:
    modsumm.append((perm[i]+array[i])%n)
    i+=1
  return modsumm
# makes list of modulo summs


def alpha():
  multiset={}
  for value in array:
    multiset[value]=modsumm.count(value)
  alpha={power: 0 for power in range(n+1)}
  for key in array:
    for power in range(n+1):
        if multiset[key]==power:
            alpha[power]+=1
  return alpha
# returns list of dictionaries (power: number of values
```

```
with this power)

def unique_alphas(alphas):
  unique_alphas=[]
  for alpha in alphas:
    if alpha not in unique_alphas:
      unique_alphas.append(alpha)
  return unique_alphas

def alpha_tupple(alpha):
  keys = range(n+1)
  alpha_list = []
  for key in keys:
    alpha_list.append(alpha[key])
  alpha_tupple = tuple(alpha_list)
  return alpha_tupple
# returns tupple of alphas <a0, a1, ...>

def file_print(perms, unique_alphas, alphas):

  headers = ["Alpha tupple".center(n*4),
  "Number of permutations".center(25)]
  f = open('{}.txt'.format(n), 'w')
  for header in headers:
    f.write(header)
  f.write('\n')

  for alpha in unique_alphas:
    f.write(str(alpha_tupple(alpha)).center(n*4))
    count = 0
```

```
    for index in range(len(perms)):
      if alphas[index]== alpha:
        count+=1
    f.write(str(count).center(25))
    f.write('\n')

  f.close()

n = 11
array = range(n)
perms = list(itertools.permutations(array))
alphas=[]
for perm in perms:
  modsumm = res(perm)
  alphas.append(alpha())
unique_alphas = unique_alphas(alphas)
file_print(perms, unique_alphas, alphas)
```

.

A.3 Код програми для генерування N випадкових перестановок і виділення класів

```
from __future__ import division
import itertools
import random
from collections import OrderedDict

import math
from numpy import *
from scipy import stats, special

def random_perm():
  perm = list(array)
  i=0
  while i<n-1:
    j=random.randint(i, n)
    perm[i], perm[j] = perm[j], perm[i]
    i += 1
  return perm

def modsumm(perm):
  modsumm=[]
  i=0
  while i<n:
    modsumm.append((perm[i]+array[i])%n)
```

```
    i+=1
  return modsumm
# makes list of modulo summs

def alpha(modsumm):
  multiset={}
  for value in array:
    multiset[value]=modsumm.count(value)
  alpha={power: 0 for power in range(n+1)}
  for key in array:
    for power in range(n+1):
        if multiset[key]==power:
            alpha[power]+=1
  return alpha
# returns list of dictionaries {power: number of values
with this power}

def alpha_tupple(alpha):
  keys = range(n+1)
  alpha_list = []
  for key in keys:
    alpha_list.append(alpha[key])
  alpha_tupple = tuple(alpha_list)
  return alpha_tupple
# returns tupple of alphas <a0, a1, ...>

def file_print():
  headers = ["Alpha tupple".center(int(round(n*3.5))),
             "Num of perms".center(15),
             "Confidence interval (1)".center(28),
             "Confidence interval (2)".center(28),
```

```
                "Confidence interval (3)".center(28),
                "Confidence interval (4)".center(28)]
    f = open('intervals_{}.txt'.format(n), 'w')
    for header in headers:
      f.write(header)
    f.write('\n')
    for item in intervals:
      f.write((str(item[0])).center(int(round(n*3.5))))
      f.write((str(int(item[1]*N))).center(15))
      f.write((str(item[2])).center(28))
      f.write((str(item[3])).center(28))
      f.write((str(item[4])).center(28))
      f.write('\n')
    f.close()

def confidence_intervals():
    confidence=0.05
    intervals = []
    M = math.factorial(n)
    for item in sorted_alphas_count:
        p = sorted_alphas_count[item]
        intervals.append([
        item, p,
        [int(M*stats.chi2.ppf(confidence, 2*p*N)/
        /(2*N-N*p+1+stats.chi2.ppf(confidence, 2*p*N)/2)),
        int(math.ceil(M*stats.chi2.ppf(1-confidence, 2*p*N+2)/
        /(2*N-N*p+stats.chi2.ppf(1-confidence, 2*p*N+2)/2)))],
        [int(M*p-M*math.sqrt(p*(1-p)/N)*
        *stats.norm.ppf(1-confidence/2)),
        int(math.ceil(M*p + M*math.sqrt(p*(1-p)/N)*
        *stats.norm.ppf(1-confidence/2)))],
```

```
        [int(stats.chi2.ppf(confidence/2, 2*p*N) *M/(2*N)),
        int(math.ceil(stats.chi2.ppf(1-confidence/2, 2*p*N+2)*
        *M/(2*N)))],
        [int(M/N*stats.binom.interval(1-confidence, N, p)[0]),
        int(math.ceil(M/
        /N*stats.binom.interval(1-confidence, N, p)[1]))]
                        ])

    return intervals


n = 55
N = 10**7

array = range(n)
alphas_count = {}
for l in range(N):
  random_per = random_perm()
  modsum = modsumm(random_per)
  alph = alpha(modsum)
  alpha_tupp = alpha_tupple(alph)
  if alpha_tupp in alphas_count:
    alphas_count[alpha_tupp] += 1
  else:
    alphas_count[alpha_tupp] = 1
sorted_alphas_count = OrderedDict(sorted(alphas_count.items(),
key=lambda x: x[1], reverse=True))
for key in sorted_alphas_count:
  sorted_alphas_count[key] /= N
intervals = confidence_intervals()
file_print()
```